PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification: C12N 15/10, A01H 5/00, C12N 5/10, C12N 15/60, C12Q 1/527, C12Q 1/68

A1

(11) International Publication Number: WO 00/28017

(43) International Publication Date: 18 May 2000 (18.05.2000)

(21) International Application Number: PCT/US99/26771

(22) International Filing Date: 09 November 1999 (09.11.1999)

(30) Priority Data:
60/107,757 10 November 1998 (10.11.1998) US

Published

(60) Parent Application or Grant
MAXYGEN, INC. [/]; (). STEMMER, Willem, P., C. [/]; (). SUBRAMANIAN, Venkiteswaran [/]; (). STEMMER, Willem, P., C. [/]; (). SUBRAMANIAN, Venkiteswaran [/]; (). QUINE, Jonathan, Alan ; ().

(54) Title: MODIFIED PHOSPHOENOLPYRUVATE CARBOXYLASE FOR IMPROVEMENT AND OPTIMIZATION OF PLANT PHENOTYPES

(54) Titre: PHOSPHOENOLPYRUVATE CARBOXYLASE MODIFIEE PERMETTANT D'AMELIORER ET D'OPTIMISER DES PHENOTYPES VEGETAUX

(57) Abstract

The invention provides methods and compositions relating to sequence-shuffled variants of PEP carboxylase in plants and microorganisms.

(57) Abrégé

Cette invention a trait à des méthodes et aux compositions correspondantes en rapport avec des variants à séquence remaniée de phosphoenolpyruvate (PEP) carboxylase dans des végétaux et des micro-organismes.

ATTORNEY DOCKET NUMBER:10424-003-999
SERIAL NUMBER: 09/920,118
REFERENCE: **B02**

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification: C12N 15/10, A01H 5/00, C12N 5/10, C12N 15/60, C12Q 1/527, C12Q 1/68 — A1

(11) International Publication Number: WO 00/28017

(43) International Publication Date: 18 May 2000 (18.05.2000)

(21) International Application Number: PCT/US99/26771

(22) International Filing Date: 09 November 1999 (09.11.1999)

(30) Priority Data:
60/107,757 10 November 1998 (10.11.1998) US

Published

(60) Parent Application or Grant
MAXYGEN, INC. [/]; (). STEMMER, Willem, P., C. [/]; (). SUBRAMANIAN, Venkiteswaran [/]; (). STEMMER, Willem, P., C. [/]; (). SUBRAMANIAN, Venkiteswaran [/]; (). QUINE, Jonathan, Alan ; ().

(54) Title: MODIFIED PHOSPHOENOLPYRUVATE CARBOXYLASE FOR IMPROVEMENT AND OPTIMIZATION OF PLANT PHENOTYPES

(54) Titre: PHOSPHOENOLPYRUVATE CARBOXYLASE MODIFIEE PERMETTANT D'AMELIORER ET D'OPTIMISER DES PHENOTYPES VEGETAUX

(57) Abstract

The invention provides methods and compositions relating to sequence-shuffled variants of PEP carboxylase in plants and microorganisms.

(57) Abrégé

Cette invention a trait à des méthodes et aux compositions correspondantes en rapport avec des variants à séquence remaniée de phosphoenolpyruvate (PEP) carboxylase dans des végétaux et des micro-organismes.

PCT
WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 15/10, C12Q 1/527, 1/68, C12N 5/10, A01H 5/00, C12N 15/60 — A1

(11) International Publication Number: WO 00/28017

(43) International Publication Date: 18 May 2000 (18.05.00)

(21) International Application Number: PCT/US99/26771

(22) International Filing Date: 9 November 1999 (09.11.99)

(30) Priority Data:
60/107,757 10 November 1998 (10.11.98) US

(71) Applicant (*for all designated States except US*): MAXYGEN, INC. [US/US]; 515 Galveston Drive, Redwood City, CA 94063 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): STEMMER, Willem, P., C. [NL/US]; 108 Kathy Court, Los Gatos, CA 95032 (US). SUBRAMANIAN, Venkiteswaran [US/US]; 3980 Corte Mar De Hierba, San Diego, CA 92130 (US).

(74) Agents: QUINE, Jonathan, Alan; The Law Offices of Jonathan Alan Quine, P.O. Box 458, Alameda, CA 94501 (US) et al.

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: MODIFIED PHOSPHOENOLPYRUVATE CARBOXYLASE FOR IMPROVEMENT AND OPTIMIZATION OF PLANT PHENOTYPES

(57) Abstract

The invention provides methods and compositions relating to sequence-shuffled variants of PEP carboxylase in plants and microorganisms.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

Modified Phosphoenolpyruvate Carboxylase for Improvement and Optimization of Plant Phenotypes

CROSS REFERENCE TO RELATED APPLICATIONS

This application is a non-provisional of and claims priority to "MODIFIED PHOSPHOENOLPYRUVATE CARBOXYLASE FOR IMPROVEMENT AND OPTIMIZATION OF PLANT PHENOTYPES" USSN 60/107,757 by Willem P.C. Stemmer and Venkitswaran Subramanian, filed November 10, 1998.

FIELD OF THE INVENTION

The invention relates to methods and compositions for generating, modifying, adapting, and optimizing polynucleotide sequences that encode proteins having PEPC enzyme activities which are useful for introduction into plant species, and other hosts, and related aspects.

BACKGROUND

Genetic Engineering of Plants

Genetic engineering of agricultural organisms dates back thousands of years to the dawn of agriculture. The hand of man has selected the agricultural organisms having the phenotypic traits that were deemed desirable, which desired phenotypic traits have often been taste, high yield, caloric value, ease of propagation, resistance to pests and disease, and appearance. Classical breeding methods to select for germplasm encoding desirable agricultural traits had been a standard practice of the world's farmers long before Gregor Mendel and others identified the basic rules of segregation and selection. For the most part, the fundamental process underlying the generation and selection of desired traits was the natural mutation frequency and recombination rates of the organisms, which are quite slow compared to the human lifespan and make it difficult to use conventional methods of breeding to rapidly obtain or optimize desired traits in an organism.

The very recent advent of non-classical, or recombinant genetic engineering techniques has provided a new means to expedite the generation of

agricultural organisms having desired traits that provide an economic, ecological, nutritional, or aesthetic benefit. To date, most recombinant approaches have involved transferring a novel or modified gene into the germline of an organism to effect its expression or to inhibit the expression of the endogenous homologue gene in the organism's native genome. However, the currently used recombinant techniques are generally unsuited for substantially increasing the rate at which a novel or improved phenotypic trait can be evolved. Essentially all recombinant genes in use today for agriculture are obtained from the germplasm of existing plant and microbial specimens, which have naturally evolved coordinately with constraints related to other aspects of the organism's evolution and typically are not optimized for the desired phenotype(s). The sequence diversity available is limited by the natural genetic variability within the existing specimen gene pool, although crude mutagenic approaches have been used to add to the natural variability in the gene pool.

Unfortunately, the induction of mutations to generate diversity often requires chemical mutagenesis, radiation mutagenesis, tissue culture techniques, or mutagenic genetic stocks. These methods provide means for increasing genetic variability in the desired genes, but frequently produce deleterious mutations in many other genes. These other traits may be removed, in some instances, by further genetic manipulation (e.g., backcrossing), but such work is generally both expensive and time consuming. For example, in the flower business, the properties of stem strength and length, disease resistance and maintaining quality are important, but often initially compromised in the mutagenesis process.

Phosphoenolpyruvate Carboxylase

Phosphoenolpyruvate (PEP) carboxylase (PEPC; EC 4.1.1.31) is a key enzyme of photosynthesis in those plant species exhibiting the C4 or CAM pathway for CO_2 fixation. The principal substrate of PEPC is the free form of PEP. PEPC catalyzes the conversion of PEP and bicarbonate to oxalacetic acid inorganic phosphate (Pi). This reaction is the first step of a metabolic route known as the C4 dicarboxylic acid pathway, which minimizes losses of energy produced by photorespiration. PEPC is present in plants, algae, cyanobacteria, and bacteria; the enzymatic properties differ based on the source.

The primary structures of PEPC from *E. coli*, *Anabaena variabilis*, and maize, among others, have been deduced from cDNA sequences and are available in the literature and GenBank. The homology found in the C-terminal half of the protein are consistent with the C-terminal half containing a catalytic domain, and the sequence between residues 603 to 616 of the *Zea mays* PEPC enzyme (-FHGRGGSIGRGGAP-) are highly conserved among taxonomic species and seem to be unique to PEPC. PEPC is a homomultimer, typically a homotetramer or homodimer, and is extrachloroplastic and located in the cytosol of the mesophyll leaves of C4 and CAM plants. Besides the C4-specific PEPC, other isozyme forms of the enzyme occur in C3 plants or etiolated C4 leaves.

PEPC from C4 plants is activated by glucose 6-phosphate (G6P), which induces an increase in Vmax and in substrate affinity for binding PEP. A metabolite, L-malate, which is an intermediate product of the carboxylation reaction, is an inhibitor of PEPC activity. It shows a cooperative effect and seems to interact with PEPC at different sites, producing noncompetitive or competitive inhibition depending on pH and concentration. G6P produces a decrease in the inhibitory effect of malate. In addition, oxaloacetate, aspartate, and certain flavenoids have been shown to inhibit PEPC (Pairoba et al. (1996) Biosci. Biotech. Biochem. 60: 779; O'Leary M (1982) Ann. Rev. Plant Physiol. 33:297). Variation in pH also controls PEPC activity: the affinity for the PEPC cofactor Mg^{+2} increases sharply between pH 7 and pH8, and the effects of the activator G6P and the inhibitor malate are more pronounced at pH 7, decreasing with increasing pH. Feedback inhibition of PEPC occurs by two distinct yet coupled mechanisms: inhibition by malate itself and enhancement of its inhibitory effect by decrease in pH which can be a consequence of malate production from PEPC activity. Another mechanism of regulating PEPC activity is post-translational modification; the interconversion of night (active) and day (inactive) forms of PEPC is mediated by phosphorylation of serine and/or threonine residues of PEPC. A variety of PEPC inhibitors have been catalogued (Devi et al. (1992) J. Plant Biochem. Biotech. 1: 73). Illumination induces a light-activable net serine dephosphorylation such that the day form is substantially inactive. As PEPC is a key control point for accomplishing the primary carboxylation of PEP,

a major component of CO_2 fixation in C4 and CAM plants, it would be desirable to have a method for producing PEPC encoding sequences and novel PEPC proteins wherein the enzymatic activity of PEPC has (1) an decreased Km for substrate, (2) a decreased Km for activator, (3) a constitutive PEPC activity in the absence of activator which is higher than naturally occurring PEPC in the absence of activator, (4) an increased Km for one or more inhibitors, (5) a desensitization to one or more inhibitors, (6) and/or (6) a higher PEPC activity in the "day form" PEPC during illumination than in a naturally-occurring PEPC "day form" under comparable illumination, or the like. Plants and other photosynthetic organisms having such enhanced PEPC encoding polynucleotides proteins would have increased net CO_2 fixation.

As noted, the advent of recombinant DNA technology has provided agriculturists with additional means of modifying plant genomes. While certainly practical in some areas, to date genetic engineering methods have had limited success in transferring or modifying important biosynthetic or other pathways in photosynthetic organisms and bacteria. The creation of plants and other photosynthetic organisms having improved PEPC biosynthetic pathways can provide increased yields of certain types of starchy foodstuffs, enhanced biomass energy sources, and may alter the types and amounts of nutrients present in certain foodstuffs, among other desirable phenotypes.

Thus, there exists a need for improved methods for producing plants and agricultural photosynthetic microbes with an improved PEPC enzyme. In particular, these methods should provide general means for producing novel PEPC enzymes, including increasing the diversity of the PEPC gene pool and the rate at which genetic sequences encoding one or more PEPC having desired properties are evolved. It is particularly desirable to have methods which are suitable for rapid evolution of genetic sequences to function in one or more plant species and confer an improved PEPC phenotype (e.g., reduced sensitivity to inhibitors (e.g., malate, pH, etc.), reduced dependence on activators (e.g., G6P, serine/threonine), improved catalytic efficiency via increasing Vmax and/or increasing the apparent affinity of substrates for the enzyme, and/or relieving a requirement for allosteric activation

(e.g., phosphorylation) or inhibition by allosteric repression, as well as plants which express the novel PEPC genetic sequence(s).

The present invention meets these and other needs and provides such improvements and opportunities.

The references discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the inventors are not entitled to antedate such disclosure by virtue of prior invention. All publications cited are incorporated herein by reference, whether specifically noted as such or not.

SUMMARY OF THE INVENTION

In a broad general aspect, the present invention provides a method for rapid evolution of polynucleotide sequences encoding a PEPC enzyme, that, when transferred into an appropriate plant cell, or photosynthetic microbial host and expressed therein, confers an enhanced metabolic phenotype to the host to increase carbon fixation ratio and/or rate, or to increase the accumulation or depletion of certain metabolites and energy storage sinks. In general, polynucleotide sequence shuffling and phenotype selection, such as detection of a parameter of PEPC enzyme activity, is employed recursively to generate polynucleotide sequences which encode novel proteins having desirable PEPC enzymatic catalytic function(s), regulatory function(s), and related enzymatic and physicochemical properties. Although the method is believed broadly applicable to evolving biosynthetic enzymes having desired properties, the invention is described principally with reference to the metabolic enzyme activities of plants and/or photosynthetic microbes and/or bacteria, defined as PEPC, or an isozyme thereof, including, respectively, plant and algal as well as bacterial forms.

PEPC Embodiment - Lowered Km for substrate

The invention provides an isolated polynucleotide encoding an enhanced PEPC protein having PEPC catalytic activity wherein the Km for a substrate (PEP, bicarbonate) is significantly lower than in a protein encoded by a parental polynucleotide encoding a naturally-occurring PEPC enzyme. Typically, the Km for substrate will be at least one-half logarithm unit lower than the parental

sequence, preferably the Km will be at least one logarithm unit lower, and desirably the Km will be at least two logarithm units lower, or more. The isolated polynucleotide encoding an enhanced PEPC protein and in an expressible form can be transferred into a host plant, such as a crop species, wherein suitable expression of the polynucleotide in the host plant will result in improved carbon fixation biosynthesis efficiency as compared to the naturally-occurring host plant species, usually under certain conditions. The isolated polynucleotide can encode a PEPC, such as a bacterial form, or may encode a PEPC enzyme such as that found in green algae, and higher plants. The isolated polynucleotide can comprise a substantially full-length or full-length coding sequence substantially identical to a naturally occurring PEPC gene and/or an isozyme thereof, typically comprising a shuffled PEPC gene.

In a variation, the invention provides a polynucleotide comprising: (1) a sequence encoding a shuffled PEPC gene operably linked to a transcriptional regulatory sequence functional in a host cell, and further linked to (2) a selectable marker gene which affords a means of selection when expressed in host cells.

In a variation, the invention provides a polynucleotide comprising: (1) a sequence encoding a shuffled PEPC gene having at least 95 percent sequence identity to a PEPC encoding sequence in the genome of a naturally-occurring plant, operably linked to a transcriptional regulatory sequence functional in a host cell, and further linked to (2) a selectable marker gene which affords a means of selection when expressed in host cells.

In a variation, the invention provides a polynucleotide comprising: (1) a sequence encoding a shuffled PEPC gene operably linked to a transcriptional regulatory sequence functional in a host cell, (2) a sequence encoding a shuffled Rubisco gene operably linked to a transcriptional regulatory sequence functional in the host cell and, optionally, further linked to (3) a selectable marker gene which affords a means of selection when expressed in host cells.

In a variation, the invention provides an isolated polynucleotide encoding an enhanced PEPC protein having PEPC catalytic activity wherein the Km for a substrate is significantly higher than a protein encoded by a parental polynucleotide encoding a naturally-occurring PEPC enzyme. In an aspect, the

enhanced PEPC protein is often catalytically active in the cytosol of cells of higher plants, particularly plants of agronomic importance. In an aspect, the enhanced PEPC protein is at least 90 percent sequence identical to a naturally occurring PEPC protein encoded by a genome of a plant or algae.

In a variation, the invention provides an isolated polynucleotide encoding an enhanced PEPC protein having PEPC catalytic activity wherein the Km (Ki) for an inhibitor (e.g., L-malate, aspartate, metabolic effectors), especially at pH levels below 8.0, is significantly higher than a protein encoded by a parental polynucleotide encoding a naturally-occurring PEPC enzyme. In such embodiments, the concentration of inhibitor required to produce half-maximal inhibition of catalysis is typically at least one-half logarithm unit higher than a parental PEPC, often at least one log unit or more higher.

In a variation, the invention provides an isolated polynucleotide encoding an enhanced PEPC protein having PEPC catalytic activity wherein the Km for an activator (e.g., glucose 6-phosphate, G6P; triose phosphate) is significantly lower than in a protein encoded by a parental polynucleotide encoding a naturally-occurring PEPC enzyme. In such embodiments, the concentration of activator required to produce half-maximal activation of catalysis is typically at least one-half logarithm unit lower than a parental PEPC, often at least one log unit or more lower, in some embodiments at least two log units or more lower. In a variation, the shuffled PEPC protein possesses, in the substantial absence of activator, PEPC catalytic activity approximately equivalent to or greater than that of a naturally-occurring PEPC protein which is maximally stimulated with activator.

The invention provides an enhanced PEPC protein having PEPC catalytic activity wherein: (1) the Km for substrate is significantly lower than in a protein encoded by a parental polynucleotide encoding a naturally-occurring PEPC enzyme, and (2) the Km for inhibitor is significantly higher than a protein encoded by a parental polynucleotide encoding a naturally-occurring PEPC enzyme, and/or (3) the Km for activator is significantly lower than in a protein encoded by a parental polynucleotide encoding a naturally-occurring PEPC enzyme, and/or (4) the enhanced PEPC protein possesses a catalytic activity in the substantial

absence of activator and inhibitor which is at least 25 percent or more greater than a naturally-occurring PEPC that is maximally stimulated with activator in the substantial absence of inhibitor and/or (5) the PEPC activity is desensitized to pH-mediated changes in allosteric control by inhibitors and/or activators; often the naturally-occurring PEPC used for comparison is an PEPC species which has a polypeptide that has the greatest percentage sequence identity to the shuffled PEPC polypeptide.

In an aspect, the invention provides a polynucleotide sequence encoding a shuffled plant or algal PEPC, wherein the shuffled PEPC protein possesses a detectable enzymatic activity wherein: (1) the Km for substrate is significantly lower than in a protein encoded by a parental polynucleotide encoding a naturally-occurring PEPC enzyme, (2) the Km for a PEPC inhibitor is significantly higher than a PEPC protein encoded by a parental polynucleotide encoding a naturally-occurring PEPC enzyme, and/or (3) the Km for an PEPC activator is significantly lower than a protein encoded by a parental polynucleotide encoding a naturally-occurring PEPC enzyme, and/or (4) the Vmax for PEPC catalytic activity is substantially higher than the Vmax for PEPC catalytic activity of naturally-occurring PEPC under equivalent assay conditions (e.g., same concentration(s) of substrates, activators, and inhibitors, and pH) under at least one assay condition. In some embodiments, the shuffled PEPC sequences encode proteins that have an altered binding to, or allosteric interaction with, a protein kinase or protein phosphatase, such that the binding constant for an inhibitor or activator on the PEPC protein may be substantially unchanged, however the shuffled PEPC, when modified by the protein kinase or phosphorylase, results in formation of a PEPC which has: (1) reduced sensitivity to inhibitors (e.g., malate) and/or (2) enhanced sensitivity to activators (e.g., G6P) or (3) has PEPC activity which is insensitive to activator and possesses at least one PEPC catalytic activity (e.g., substrate Km^{-1} or Vmax) which is at least 25 percent greater than that of a naturally-occurring PEPC that is maximally stimulated with activator in the substantial absence of inhibitor; often the naturally-occurring PEPC used for comparison is a PEPC species which has a polypeptide that has the greatest percentage sequence identity, among the collection of then known PEPC

sequences, to the shuffled PEPC polypeptide. In some embodiments, the binding constant for an inhibitor, activator, and/or substrate will be at least one-half log unit higher or lower than an equivalent naturally occurring PEPC of greatest sequence homology (percent sequence identity) to the shufflant.

In an aspect, the invention provides an improved PEPC, or shufflant thereof, and a polynucleotide encoding same. In some embodiments, the polynucleotide will be operably linked to a transcription regulation sequence forming an expression construct, which may be linked to a selectable marker gene. In some embodiments, such a PEPC polynucleotide is present as an integrated transgene in a plant chromosome in a format for expression and processing of the enzyme. It can be desirable for such a polynucleotide transgene to be transmissible via germline transmission in a plant; in the case of PEPC gene sequences transferred to plant or algal cells, it is often accompanied by a selectable marker gene which affords a means to select for progeny which retain the transferred shuffled PEPC gene sequence. In some embodiments, the transferred shuffled PEPC gene sequence is derived by shuffling a pool of parental sequences, at least one of which encodes a bacterial PEPC. Often, the transcription control sequences comprise tissue-specific or conditional promoters to overcome possible detrimental effects of constitutive expression.

In an aspect, the invention provides an improved PEPC, or shufflant thereof, wherein the improved PEPC has at least 80 sequence identity to the polypeptide sequence of a naturally-occurring plant PEPC, and which has an enhanced PEPC enzymatic phenotype; and a polynucleotide encoding same. In some embodiments, the polynucleotide will be operably linked to a transcription regulation sequence forming an expression construct, which may be linked to a selectable marker gene. In some embodiments, such a PEPC polynucleotide is present as an integrated transgene in a plant chromosome and may be accompanied, in linked or unlinked configuration, with a Rubisco encoding polynucleotide and/or an ADPGPP encoding polynucleotide; often such Rubisco and/or ADPGPP polynucleotides encode an optimized, shuffled enzyme.

In an aspect, the invention provides a hybrid PEPC composed of a shufflant comprising a sequence of at least 25 contiguous nucleotides at least 95 percent identical to a plant PEPC gene and a sequence of at least 25 contiguous nucleotides at least 95 percent identical to a bacterial or algal PEPC gene, and a polynucleotide encoding same, and typically encoding a substantially full-length PEPC protein, usually comprising at least 90 percent of the coding sequence length, but not necessarily sequence identity, of a naturally occurring PEPC protein. In some embodiments, the polynucleotide will be operably linked to a transcription regulation sequence forming an expression construct, which may be linked to a selectable marker gene. In some embodiments, such a polynucleotide is present as an integrated transgene in a plant chromosome. It can be desirable for such a polynucleotide transgene to be transmissible via germline transmission in a plant.

The invention provides expression constructs, including bacterial plasmids, shuttle vectors, and plant transgenes, wherein the expression construct comprises a transcriptional regulatory sequence functional in plants operably linked to a polynucleotide encoding an enhanced PEPC protein. With respect to polynucleotide sequences encoding PEPC proteins, it is generally desirable to express such encoding sequences in plant cells with the expression constructs containing the necessary sequences for appropriate transcription, translation, and processing. The invention further provides plants and plant germplasm comprising said expression constructs, typically in stably integrated or other replicable form which segregates and can be stably maintained in the host organism, although in some embodiments it is desirable for commercial reasons that the expression sequence not be in the germline of sexually repoducible plants.

The invention provides a method for obtaining an isolated polynucleotide encoding an enhanced PEPC protein having PEPC catalytic activity wherein the Km for substrate is significantly lower than a protein encoded by a parental polynucleotide encoding a naturally-occurring PEPC enzyme, the method comprising: (1) recombining sequences of a plurality of parental polynucleotide species encoding at least one PEPC sequence under conditions suitable for sequence shuffling to form a resultant library of sequence-shuffled PEPC polynucleotides, (2)

transferring said library into a plurality of host cells forming a library of transformants wherein sequence-shuffled PEPC polynucleotides are expressed, (3) assaying individual or pooled transformants for PEPC catalytic activity to determine the relative or absolute Km for substrate and identifying at least one enhanced transformant that expresses a PEPC activity which has a significantly lower Km for substrate than the PEPC activity encoded by the parental sequence(s), (4) recovering the sequence-shuffled PEPC polynucleotide from at least one enhanced transformant. Optionally, the recovered sequence-shuffled PEPC polynucleotide encoding an enhanced PEPC is recursively shuffled and selected by repeating steps 1 through 4, wherein the recovered sequence-shuffled PEPC polynucleotide is used as at least one parental sequence for subsequent shuffling. If it is desired to obtain a sequence-shuffled PEPC encoding a PEPC enzyme having an increased Km for inhibitor, step 3 comprises assaying individual or pooled transformants for PEPC catalytic activity to determine the relative or absolute Km for the inhibitor and identifying at least one enhanced transformant that expresses a PEPC activity which has a significantly higher Km for inhibitor than the PEPC activity encoded by the parental sequence(s). Similarly, if it is desired to obtain a sequence-shuffled PEPC encoding a PEPC enzyme having a decreased Km for activator, step 3 comprises assaying individual or pooled transformants for PEPC catalytic activity to determine the relative or absolute Km for activator, and identifying at least one enhanced transformant that expresses a PEPC activity which has a significantly lower Km for activator than the PEPC activity encoded by the parental sequence(s).

In an aspect, the PEPC gene sequence(s) is/are obtained as an isolated polynucleotide and is shuffled by any suitable shuffling method known in the art, such as DNA fragmentation and PCR, error-prone PCR, and the like, preferably with one or more additional parental polynucleotides encoding all or a part of another PEPC species. The population of sequence-shuffled PEPC polynucleotides are each operably linked to an expression sequence and transferred into host cells, preferably host cells substantially lacking endogenous PEPC activity, wherein the sequence-shuffled PEPC polynucleotides are expressed, forming a library of sequence-shuffled PEPC transformants. A sample of individual transformants and/or their clonal

progeny are isolated into discrete reaction vessels for PEPC activity assay, or are assayed in situ in certain embodiments. For samples assayed in reaction vessels, aliquots of the samples are separated into a plurality of reaction vessels containing an approximately equimolar amount of PEPC or total protein, and each vessel is assayed for PEPC activity in the presence of a predetermined concentration of substrate which ranges from about 0.0001 times the predetermined Km for substrate of the PEPC encoded by the parental polynucleotide(s) to about 10,000 times the predetermined Km for substrate of the PEPC encoded by the parental polynucleotide(s); the plurality of reaction vessels for each shufflant sample may also contain a fixed or variable concentration of activator and/or inhibitor, or neither. From the data generated by assaying the plurality of reaction vessels containing aliquots of each transformant, a Km value and/or Vmax is calculated by conventional art-known means for the sequence-shuffled PEPC of each transformant; typically the Km and Vmax values for a specific inhibitor or activator are determined. Sequence-shuffled polynucleotides encoding PEPC proteins that have significantly decreased Km and/or Vmax values for substrate, and/or significantly increased Km values of inhibitor, and/or significantly decreased Km values for activator are selected and used as parental sequences for at least one additional round of sequence shuffling by any suitable method and selection for further optimization of the desired PEPC phenotype. The shuffling and selection process is performed iteratively until sequence shuffled polynucleotides encoding at least one PEPC enzyme having a desired PEPC enzymatic phenotype is obtained, or until the optimization to reduce the relevant Km (or increase Vmax) has plateaued and no further improvement is seen in subsequent rounds of shuffling and selection.

In a variation, the sequence-shuffled polynucleotides operably linked to an expression sequence is also linked, in polynucleotide linkage, to an expression cassette encoding a selectable marker gene. Transformants are propagated on a selective medium to ensure that transformants which are assayed for PEPC activity contain a sequence-shuffled PEPC encoding sequence in expressible form.

In a variation, the above-described method is modified such that PEPC activity is assayed in the presence of varying concentrations of inhibitor and the Km

for inhibitor is determined. Each vessel containing an aliquot of a transformant is assayed for PEPC activity in the presence of a predetermined concentration of inhibitor which ranges from about 0.0001 times the predetermined Km for inhibitor of the PEPC encoded by the parental polynucleotide(s) to about 10,000 times the predetermined Km for inhibitor of the PEPC encoded by the parental polynucleotide(s). From the data generated by assaying the plurality of reaction vessels containing aliquots of each transformant, a Km value is calculated by conventional art-known means for the sequence-shuffled PEPC of each transformant. Sequence-shuffled polynucleotides encoding PEPC proteins that have significantly increased Km values for inhibitor are selected and used as parental sequences for at least one additional round of sequence shuffling by any suitable method and selection for increased Km values for inhibitor. The shuffling and selection process is performed iteratively until sequence shuffled polynucleotides encoding at least one PEPC enzyme having a desired Km value is obtained, or until the optimization to increase the Km has plateaued and no further improvement is seen in subsequent rounds of shuffling and selection.

In a variation, the above-described method is modified such that PEPC activity is assayed in the presence of varying concentrations of activator and the Km for activator is determined. Each vessel containing an aliquot of a transformant is assayed for PEPC activity in the presence of a predetermined concentration of activator which ranges from about 0.0001 times the predetermined Km for activator of the PEPC encoded by the parental polynucleotide(s) to about 10,000 times the predetermined Km for activator of the PEPC encoded by the parental polynucleotide(s). From the data generated by assaying the plurality of reaction vessels containing aliquots of each transformant, a Km value is calculated by conventional art-known means for the sequence-shuffled PEPC of each transformant. Sequence-shuffled polynucleotides encoding PEPC proteins that have significantly decreased Km values for activator are selected and used as parental sequences for at least one additional round of sequence shuffling by any suitable method and selection for decreased Km values for activator. The shuffling and selection process is performed iteratively until sequence shuffled polynucleotides encoding at least one

PEPC enzyme having a desired Km value is obtained, or until the optimization to increase the Km has plateaued and no further improvement is seen in subsequent rounds of shuffling and selection.

In a variation, the method comprises conducting biochemical assays on sample aliquots of transformants to determine PEPC enzyme activity so as to establish the ratio of the Km for activator to the Km for inhibitor for individual transformants. Sequence-shuffled polynucleotides encoding PEPC are obtained from transformants exhibiting a decrease in said ratio as compared to the ratio in PEPC produced from the parental encoding polynucleotide(s) to provide selected sequence-shuffled PEPC polynucleotides which can be used as parental sequences for at least one additional round of sequence shuffling by any suitable method and selection for a decreased ratio of Km(activator) to Km(inhibitor). The shuffling and selection process is performed iteratively until sequence shuffled polynucleotides encoding at least one PEPC enzyme having a desired Km ratio is obtained, or until the optimization to decrease the Km ratio has plateaued and no further improvement is seen in subsequent rounds of shuffling and selection.

In a variation, the method comprises conducting biochemical assays on sample aliquots of transformants to determine the pH profile of PEPC enzyme activity and the pH sensitivity of activator and inhibitor effects. A pH desensitized PEPC exhibits PEPC activity such that an increase in pH from approximately 7.0 to 8.0 produces: (1) a decrease in the Ki of malate or other inhibitor of less than one half of the decrease seen in parental PEPC enzyme under identical conditions, and/or (2) an increase in Km of activator of less than one half of the increase seen in parental PEPC enzyme under identical conditions. Sequence-shuffled polynucleotides encoding PEPC are obtained from transformants exhibiting a decrease in pH effect as compared to the produced from the parental encoding polynucleotide(s) to provide selected sequence-shuffled PEPC polynucleotides which can be used as parental sequences for at least one additional round of sequence shuffling by any suitable method and selection for a decreased ratio of Km(activator) to Km(inhibitor). The shuffling and selection process is performed iteratively until sequence shuffled polynucleotides encoding at least one PEPC enzyme having a desired Km ratio is obtained, or until the

optimization to decrease the Km ratio has plateaued and no further improvement is seen in subsequent rounds of shuffling and selection.

In an embodiment of the method, the host cell for transformation with sequence-shuffled polynucleotides encoding PEPC is a bacterial mutant which lacks a functional PEPC protein, such as *E. coli* mutant or an equivalent.

In an embodiment of the method, polynucleotides encoding naturally-occurring PEPC protein sequences of a plurality of species of photosynthetic prokaryotes and/or algae and/or higher plants are shuffled by a suitable shuffling method to generate a shuffled PEPC polynucleotide library, wherein each shuffled PEPC encoding sequence is operably linked to an expression sequence, and which may optionally comprise a linked selectable marker gene cassette. Said library is transformed into a host cell population to form a transformed host cell library. The transformed host cell library is propagated on growth medium, which may contain a selection agent to ensure retention of a linked selectable marker gene. Transformed host cells which are screened for under the most stringent conditions are isolated individually or in pools, and the sequence-shuffled polynucleotide sequences encoding PEPC are recovered, and optionally subjected to at least one subsequent iteration of shuffling and selection on growth medium and PEPC activity screening. Optionally or in addition, transformants are assayed for inhibitor-resistant PEPC activity and/or high activity PEPC in absence of activator. The recovered sequence-shuffled PEPC polynucleotide(s) encode(s) an enhanced PEPC protein.

The invention provides a plant cell protoplast and clonal progeny thereof containing a sequence-shuffled polynucleotide encoding a PEPC which is not encoded by the naturally occurring genome of the plant cell protoplast. The invention also provides a collection of plant cell protoplasts transformed with a library of sequence-shuffled PEPC polynucleotides in expressible form.

The invention also provides a regenerated plant containing at least one species of replicable or integrated polynucleotide comprising a sequence-shuffled portion and encoding a PEPC polypeptide. The invention provides a method variation wherein at least one round of phenotype selection is performed on regenerated plants derived from protoplasts transformed with sequence-shuffled PEPC library members.

In an embodiment, the phenotype selection comprises a determination, either directly or by proxy, of carbon fixation via the PEPC reaction.

The invention provides species-specific PEPC shuffling, wherein a transformed plant cell or adult plant or reproductive structure comprises a polynucleotide encoding a shuffled PEPC that is at least 95 percent sequence identical to the corresponding PEPC encoded by an untransformed naturally-occurring genome of the same taxonomic species of plant cell or adult plant. Typically, the shuffled PEPC results from shuffling of one or more alleles encoding the PEPC in the taxonomic species genome, optionally including mutagenesis in one or more of the iterative shuffling and selection cycles. The species-specific PEPC shuffling may include shuffling a polynucleotide encoding a full-length PEPC of a first taxonomic species under conditions whereby PEPC sequences of a second taxonomic species (or collection of species) are shuffled in at a low prevalence, such that the resultant population of shufflant polynucleotides contains, on average, shuffled polynucleotides composed of at least about 95 percent sequence encoding the first taxonomic species PEPC and less than about 5 percent sequence encoding the second taxonomic species (or collection of species) PEPC. The species-specific shufflants are thus highly biased towards identity with the first taxonomic species and shufflants which are selected for the desired PEPC phenotype are transferred back into the first taxonoic species for expression and regeneration of adult plants and germplasm. Optionally, selected shufflants are backcrossed against the naturally occurring PEPC encoding sequences of the first taxonomic species to remove non-essential sequence alterations and harmonize the final shufflant sequence to the naturally-occurring PEPC sequence of the first taxonomic species.

A variation of the method includes adapting a bacterial or algal PEPC for optimal function in a plant cell, or adult vegetative plant. This variation comprises recursive shuffling and selection of a library of bacterial or algal PEPC encoding sequences in a plant cell of the taxonomic species of plant for which the bacterial or algal PEPC is being adapted to function in an adult plant. This variation can include not only selecting for a desired PEPC enzymatic phenotype, but also selecting for appropriate function of a operably linked transcriptional control

sequence in conjunction with PEPC function. This variation can employ host cells which are regenerable post-transformation, and selection of adult plants for enhanced carbon fixation via PEPC; recovery of the encoding PEPC shufflants (and optionally the linked transcriptional control sequences), and at least one cycle of recursive shuffling and selection to evolve a bacterial or algal PEPC, and optionally a transcriptional control sequence, optimized for function in the desired plant taxonomic species or closely related taxonomic categories.

An object of the invention is the production of higher plants which express one or more PEPC enzyme which confer an enhanced carbon fixation conversion ratio to the plants. Although the invention is described principally with respect to the use of genetic sequence shuffling to generate enhanced PEPC coding sequences, the invention also provides for the introduction of PEPC coding sequences obtained from organisms having PEPC with desirable enzymatic phenotypes, such as inhibitor-resistant PEPC from bacterial mutants, into higher plants. Thus, the invention provides a method comprising the step of introducing into a higher plant (e.g., a monocot or dicot) an expression cassette encoding a PEPC encoded by a genome of a bacterium or algae. Typically, at least a sequence encoding a substantially full-length PEPC protein of the bacterial or algal PEPC is transferred. An aspect of the invention provides C4 land plants comprising a polynucleotide sequence encoding a bacterial or algal PEPC composed in an expression cassette suitable for expression in a C4 land plant; optionally an expression cassette encoding a PEPC operably linked to regulatory sequences for expression in the nucleus of the C4 plant, e.g., in tissue such as mesophyll cells, additionally is transferred into the nucleus of the C4 plant. A C3 plant may be used in place of a C4 plant if desired. A specific embodiment comprises a regenerable protoplast of *Glycine max*, *Nicotiana tabacum*, or *Zea mays* (or other agricultural crop species amenable to regeneration from protoplasts) having a nuclear genome containing an expressible shuffled PEPC gene that is obtained from a bacterium or algae, and typically is at least 90 percent up to 99 percent sequence identical to a PEPC gene in the genome of said bacterium or algae, but is mutated in at least one codon as compared to the parental sequence. The invention also provides adult plants, cultivars, seeds, vegetative bodies, fruits,

germplasm, and reproductive cells obtained from regeneration of such transformed protoplasts.

The invention provides a kit for obtaining a polynucleotide encoding a PEPC protein having a predetermined enzymatic phenotype, the kit comprising a cell line suitable for forming transformable host cells and a collection sequence-shuffled polynucleotides formed by in vitro sequence shuffling. The kit often further comprises a transformation enhancing agent (e.g., lipofection agent, PEG, etc.) and/or a transformation device (e.g., a biolistics gene gun) and/or a plant viral vector which can infect plant cells or protoplasts thereof.

The disclosed method for providing an agricultural organism having an improved PEPC enzymatic phenotype by iterative gene shuffling and phenotype selection is a pioneering method which enables a broad range of novel and advantageous agricultural compositions, methods, kits, uses, plant cultivars, and apparatus which will be apparent to those skilled in the art in view of the present disclosure.

Other features and advantages of the invention will be apparent from the following description of the drawings, preferred embodiments of the invention, the examples, and the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. Desensitization of PEPC to activator and inhibitor. Panel A shows a diagrammatic representation of PEPC activity as a function of activator concentration for a parental wild-type PEPC (solid line), a shufflant which is partially desensitized (dotted line), and a shufflant which is fully desensitized (dashed line) to activator. Panel B shows a diagrammatic representation of PEPC activity as a function of inhibitor concentration for a parental wild-type PEPC (solid line), a shufflant which is partially desensitized (dotted line), and a shufflant which is fully desensitized (dashed line) to inhibitor.

Figure 2. Optimization by shuffling of PEPC for substrate usage and resistance to inhibition. Panel A shows a diagrammatic representation of PEPC activity as a function of substrate concentration for a parental wild-type PEPC (solid line), and a shufflant which is optimized for substrate usage(dashed line); Km for the

wildtype Km(wt) and optimized enzyme Km(opt), and Vmax for the wildtype Vmax(wt) and optimized Vmax(opt) are shown. Panel B shows a diagrammatic representation of PEPC activity as a function of inhibitor concentration for a parental wild-type PEPC (solid line), and a shufflant which is optimized for substrate usage(dashed line); Km for the wildtype Km(wt) and optimized enzyme Km(opt), and Vmax for the wildtype Vmax(wt) and optimized Vmax(opt) are shown.

DETAILED DESCRIPTION

Definitions

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described. For purposes of the present invention, the following terms are defined below.

The term "shuffling" is used herein to indicate recombination between similar but non-identical polynucleotide sequences. Generally, more than one cycle of recombination is performed in DNA shuffling methods. In some embodiments, DNA shuffling may involve crossover via nonhomologous recombination, such as via cre/lox and/or flp/frt systems and the like, such that recombination need not require substantially homologous polynucleotide sequences. In silico and oligonucleotide mediated approaches also do not require similarity/homology. Homologous and non-homologous recombination formats can be used, and, in some embodiments, can generate molecular chimeras and/or molecular hybrids of substantially dissimilar sequences. Viral recombination systems, such as template-switching and the like can also be used to generate molecular chimeras and recombined genes, or portions thereof. A general description of shuffling is provided in commonly-assigned WO98/13487 and WO98/13485, both of which are incorporated herein in their entirety by reference; in case of any conflicting description of definition between any of the incorporated documents and the text of this specification, the present specification provides the principal basis for guidance and disclosure of the present invention.

The term "related polynucleotides" means that regions or areas of the polynucleotides are identical and regions or areas of the polynucleotides are heterologous.

The term "chimeric polynucleotide" means that the polynucleotide comprises regions which are wild-type and regions which are mutated. It may also mean that the polynucleotide comprises wild-type regions from one polynucleotide and wild-type regions from another related polynucleotide.

The term "cleaving" means digesting the polynucleotide with enzymes or breaking the polynucleotide (e.g., by chemical or physical means), or generating partial length copies of a parent sequence(s) via partial PCR extension, PCR stuttering, differential fragment amplification, or other means of producing partial length copies of one or more parental sequences.

The term "population" as used herein means a collection of components such as polynucleotides, nucleic acid fragments or proteins. A "mixed population" means a collection of components which belong to the same family of nucleic acids or proteins (i.e. are related) but which differ in their sequence (i.e. are not identical) and hence in their biological activity.

The term "mutations" means changes in the sequence of a parent nucleic acid sequence (e.g., a gene or a microbial genome, transferable element, or episome) or changes in the sequence of a parent polypeptide. Such mutations may be point mutations such as transitions or transversions. The mutations may be deletions, insertions or duplications.

The term "recursive sequence recombination" as used herein refers to a method whereby a population of polynucleotide sequences are recombined with each other by any suitable recombination means (e.g., sexual PCR, homologous recombination, site-specific recombination, etc.) to generate a library of sequence-recombined species which is then screened or subjected to selection to obtain those sequence-recombined species having a desired property; the selected species are then subjected to at least one additional cycle of recombination with themselves and/or with other polynucleotide species and at subsequent selection or screening for the desired property.

The term "amplification" means that the number of copies of a nucleic acid fragment is increased.

The term "naturally-occurring" as used herein as applied to an object refers to the fact that an object can be found in nature. For example, a polypeptide or polynucleotide sequence that is present in an organism that can be isolated from a source in nature and which has not been intentionally modified by man in the laboratory is naturally-occurring. As used herein, laboratory strains and established cultivars of plants which may have been selectively bred according to classical genetics are considered naturally-occurring. As used herein, naturally-occurring polynucleotide and polypeptide sequences are those sequences, including natural variants thereof, which can be found in a source in nature, or which are sufficiently similar to known natural sequences that a skilled artisan would recognize that the sequence could have arisen by natural mutation and recombination processes.

As used herein "predetermined" means that the cell type, non-human animal, or virus may be selected at the discretion of the practitioner on the basis of a known phenotype.

As used herein, "linked" means in polynucleotide linkage (i.e., phosphodiester linkage). "Unlinked" means not linked to another polynucleotide sequence; hence, two sequences are unlinked if each sequence has a free 5' terminus and a free 3' terminus.

As used herein, the term "operably linked" refers to a linkage of polynucleotide elements in a functional relationship. A nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For instance, a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the coding sequence. Operably linked means that the DNA sequences being linked are typically contiguous and, where necessary to join two protein coding regions, contiguous and in reading frame. However, since enhancers generally function when separated from the promoter by several kilobases and intronic sequences may be of variable lengths, some polynucleotide elements may be operably linked but not contiguous. A structural gene (e.g., a *PEPC* gene) which is operably linked to a polynucleotide sequence corresponding to a transcriptional

regulatory sequence of an endogenous gene is generally expressed in substantially the same temporal and cell type-specific pattern as is the naturally-occurring gene.

As used herein, the terms "expression cassette" refers to a polynucleotide comprising a promoter sequence and, optionally, an enhancer and/or silencer element(s), operably linked to a structural sequence, such as a cDNA sequence or genomic DNA sequence. In some embodiments, an expression cassette may also include polyadenylation site sequences to ensure polyadenylation of transcripts. When an expression cassette is transferred into a suitable host cell, the structural sequence is transcribed from the expression cassette promoter, and a translatabble message is generated, either directly or following appropriate RNA splicing. Typically, an expression cassette comprises: (1) a promoter, such as a CaMV 35S promoter, a NOS promoter or a *rbcS* promoter, or other suitable promoter known in the art, (2) a cloned polynucleotide sequence, such as a cDNA or genomic fragment ligated to the promoter in sense orientation so that transcription from the promoter will produce a RNA that encodes a functional protein, and (3) a polyadenylation sequence. For example and not limitation, an expression cassette of the invention may comprise the cDNA expression cloning vectors, pCD and λNMT (Okayama H and Berg P (1983) Mol. Cell. Biol. 3: 280; Okayama H and Berg P (1985) Mol. Cell. Biol. 5: 1136, incorporated herein by reference).

As used herein, the term "transcriptional unit" or "transcriptional complex" refers to a polynucleotide sequence that comprises a structural gene (exons), a cis-acting linked promoter and other cis-acting sequences necessary for efficient transcription of the structural sequences, distal regulatory elements necessary for appropriate tissue-specific and developmental transcription of the structural sequences, and additional cis sequences important for efficient transcription and translation (e.g., polyadenylation site, mRNA stability controlling sequences).

As used herein, the term "transcription regulatory region" refers to a DNA sequence comprising a functional promoter and any associated transcription elements (e.g., enhancer, CCAAT box, TATA box, LRF, ethanol-inducible element, etc.) that are essential for transcription of a polynucleotide sequence that is operably linked to the transcription regulatory region.

As used herein, the term "xenogeneic" is defined in relation to a recipient genome, host cell, or organism and means that an amino acid sequence or polynucleotide sequence is not encoded by or present in, respectively, the naturally-occurring genome of the recipient genome, host cell, or organism. Xenogenic DNA sequences are foreign DNA sequences. Further, a nucleic acid sequence that has been substantially mutated (e.g., by site directed mutagenesis) is xenogeneic with respect to the genome from which the sequence was originally derived, if the mutated sequence does not naturally occur in the genome.

The term "corresponds to" is used herein to mean that a polynucleotide sequence is homologous (i.e., identical) to all or a portion of a reference polynucleotide sequence, or that a polypeptide sequence is identical to a reference polypeptide sequence. In contradistinction, the term "complementary to" is used herein to mean that the complementary sequence is homologous to all or a portion of a reference polynucleotide sequence. For illustration, the nucleotide sequence "5'-TATAC" corresponds to a reference sequence "5'-TATAC" and is complementary to a reference sequence "5'-GTATA".

The following terms are used to describe the sequence relationships between two or more polynucleotides: "reference sequence", "comparison window", "sequence identity", "percentage of sequence identity", and "substantial identity". A "reference sequence" is a defined sequence used as a basis for a sequence comparison; a reference sequence may be a subset of a larger sequence, for example, as a segment of a full-length viral gene or virus genome. Generally, a reference sequence is at least 20 nucleotides in length, frequently at least 25 nucleotides in length, and often at least 50 nucleotides in length. Since two polynucleotides may each comprise (1) a sequence (i.e., a portion of the complete polynucleotide sequence) that is similar between the two polynucleotides, and (2) a sequence that is divergent between the two polynucleotides, sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of sequence similarity.

A "comparison window", as used herein, refers to a conceptual segment of at least 25 contiguous nucleotide positions wherein a polynucleotide sequence may be compared to a reference sequence of at least 25 contiguous nucleotides and wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) of 20 percent or less as compared to the reference sequence (which for comparative purposes in this manner does not comprise additions or deletions) for optimal alignment of the two sequences. Optimal alignment of sequences for aligning a comparison window may be conducted by the local homology algorithm of Smith and Waterman (1981) Adv. Appl. Math. 2: 482, by the homology alignment algorithm of Needleman and Wunsch (1970) J. Mol. Biol. 48: 443, by the search for similarity method of Pearson and Lipman (1988) Proc. Natl. Acad. Sci. (U.S.A.) 85: 2444, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Dr., Madison, WI), or by inspection, and the best alignment (i.e., resulting in the highest percentage of homology over the comparison window) generated by the various methods is selected.

The term "sequence identity" means that two polynucleotide sequences are identical (i.e., on a nucleotide-by-nucleotide basis) over the window of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G, U, or I) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 89 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison window of at least 20 nucleotide positions, optionally over a window of at least 30-50

nucleotides, wherein the percentage of sequence identity is calculated by comparing the reference sequence to the polynucleotide sequence that may include deletions or additions which total 20 percent or less of the reference sequence over the window of comparison. The reference sequence may be a subset of a larger sequence.

Specific hybridization is defined herein as the formation, by hydrogen bonding or nucleotide (or nucleobase) bases, of hybrids between a probe polynucleotide (e.g., a polynucleotide of the invention and a specific target polynucleotide, wherein the probe preferentially hybridizes to the specific target such that, for example, a single band corresponding to, e.g., one or more of the RNA species of the gene (or specifically cleaved or processed RNA species) can be identified on a Northern blot of RNA prepared from a suitable source. Such hybrids may be completely or only partially base-paired. Polynucleotides of the invention which specifically hybridize to viral genome sequences may be prepared on the basis of the sequence data provided herein and available in the patent applications incorporated herein and scientific and patent publications noted above, and according to methods and thermodynamic principles known in the art and described in Sambrooke et al. et al., Molecular Cloning: A Laboratory Manual, 2nd Ed., (1939), Cold Spring Harbor, N.Y.; Berger and Kimmel, Methods in Enzymology, Volume 152, Guide to Molecular Cloning Techniques (1987), Academic Press, Inc., San Diego, CA; Goodspeed et al. (1989) Gene 76: 1; Dunn et al. (1989) J. Biol. Chem. 264: 13057, and Dunn et al. (1988) J. Biol. Chem. 263: 10878, which are each incorporated herein by reference.

"Physiological conditions" as used herein refers to temperature, pH, ionic strength, viscosity, and like biochemical parameters that are compatible with a viable plant organism or agricultural microorganism (e.g., Rhizobium, Agrobacterium, etc.), and/or that typically exist intracellularly in a viable cultured plant cell, particularly conditions existing in the nucleus of said cell. In general, in vitro physiological conditions can comprise 50-200 mM NaCl or KCl, pH 6.5-8.5, 20-45EC and 0.001-10 mM divalent cation (e.g., Mg^{++}, Ca^{++}); preferably about 150 mM NaCl or KCl, pH 7.2-7.6, 5 mM divalent cation, and often include 0.01-1.0 percent nonspecific protein (e.g., BSA). A non-ionic detergent (Tween, NP-40, Triton X-

100) can often be present, usually at about 0.001 to 2%, typically 0.05-0.2% (v/v). Particular aqueous conditions may be selected by the practitioner according to conventional methods. For general guidance, the following buffered aqueous conditions may be applicable: 10-250 mM NaCl, 5-50 mM Tris HCl, pH 5-8, with optional addition of divalent cation(s), metal chelators, nonionic detergents, membrane fractions, antifoam agents, and/or scintillants.

As used herein, the terms "label" or "labeled" refer to incorporation of a detectable marker, e.g., a radiolabeled amino acid or a recoverable label (e.g. biotinyl moieties that can be recovered by avidin or streptavidin). Recoverable labels can include covalently linked polynucleobase sequences that can be recovered by hybridization to a complementary sequence polynucleotide. Various methods of labeling polypeptides, PNAs, and polynucleotides are known in the art and may be used. Examples of labels include, but are not limited to, the following: radioisotopes (e.g., ^{3}H, ^{14}C, ^{35}S, ^{125}I, ^{131}I), fluorescent or phosphorescent labels (e.g., FITC, rhodamine, lanthanide phosphors), enzymatic labels (e.g., horseradish peroxidase, β-galactosidase, luciferase, alkaline phosphatase), biotinyl groups, predetermined polypeptide epitopes recognized by a secondary reporter (e.g., leucine zipper pair sequences, binding sites for antibodies, transcriptional activator polypeptide, metal binding domains, epitope tags). In some embodiments, labels are attached by spacer arms of various lengths, e.g., to reduce potential steric hindrance.

As used herein, the term "statistically significant" means a result (i.e., an assay readout) that generally is at least two standard deviations above or below the mean of at least three separate determinations of a control assay readout and/or that is statistically significant as determined by Student's t-test or other art-accepted measure of statistical significance.

The term "transcriptional modulation" is used herein to refer to the capacity to either enhance transcription or inhibit transcription of a structural sequence linked in cis; such enhancement or inhibition may be contingent on the occurrence of a specific event, such as stimulation with an inducer and/or may only be manifest in certain cell types.

The term "agent" is used herein to denote a chemical compound, a mixture of chemical compounds, a biological macromolecule, or an extract made from biological materials such as bacteria, plants, fungi, or animal cells or tissues. Agents are evaluated for potential activity as PEPC inhibitors or allosteric effectors by inclusion in screening assays described hereinbelow.

As used herein, "substantially pure" means an object species is the predominant species present (i.e., on a molar basis it is more abundant than any other individual macromolecular species in the composition), and preferably a substantially purified fraction is a composition wherein the object species comprises at least about 50 percent (on a molar basis) of all macromolecular species present. Generally, a substantially pure composition will comprise more than about 80 to 90 percent of all macromolecular species present in the composition. Most preferably, the object species is purified to essential homogeneity (contaminant species cannot be detected in the composition by conventional detection methods) wherein the composition consists essentially of a single macromolecular species. Solvent species, small molecules (<500 Daltons), and elemental ion species are not considered macromolecular species.

As used herein, the term "optimized" is used to mean substantially improved in a desired structure or function relative to an initial starting condition, not necessarily the optimal structure or function which could be obtained if all possible combinatorial variants could be made and evaluated, a condition which is typically impractical due to the number of possible combinations and permutations in polynucleotide sequences of significant length (e.g., a complete plant gene or genome).

As used herein, "PEPC enzymatic phenotype" means an observable or otherwise detectable phenotype that can be discriminative based on PEPC function. For example and not limitation, a PEPC enzymatic phenotype can comprise an enzyme Km for a substrate, Km for an inhibitor (K_I), Km for an activator (Ka), Vmax, a turnover rate, an inhibition coefficient (Ki), or an observable or otherwise detectable trait that reports PEPC function in a cell or clonal progeny thereof, including an adult

plant or organ thereof, which otherwise lack said trait in the absence of significant PEPC function.

Description of Preferred Embodiments

The present invention provides methods, reagents, genetically modified plants, plant cells and protoplasts thereof, microbes, and polynucleotides, and compositions relating to the forced evolution of PEPC sequences to improve an enzymatic property of a PEPC protein. In an aspect, the invention provides a shuffled PEPC which is catalytically active and which exhibits an improved enzymatic profile, such as an increased Km for inhibitor, decreased Km for activator, and or a decreased Km for substrate, increased Vmax, reduced pH sensitivity, or the like.

In a broad aspect, the invention is based, in part, on a method for shuffling polynucleotide sequences that encode a PEPC enzyme. The method comprises the step of selecting at least one polynucleotide sequence that encodes a PEPC having an enhanced enzymatic phenotype and subjecting said selected polynucleotide sequence to at least one subsequent round of mutagenesis and/or sequence shuffling, and selection for the enhanced phenotype. Preferably, the method is performed recursively on a collection of selected polynucleotide sequences encoding the PEPC to iteratively provide polynucleotide sequences encoding PEPC species having the desired enhanced enzymatic phenotype.

The invention provides shuffled PEPC encoding sequences, wherein said shuffled encoding sequences comprise at least 21 contiguous nucleotides, preferably at least 30 contiguous nucleotides, or more, of a first naturally occurring PEPC gene sequence and at least 21 contiguous nucleotides, preferably at least 30 contiguous nucleotides, or more, of a second naturally occurring PEPC sequence, operably linked in reading frame to encode a PEPC which has PEPC activity and which has an enhanced PEPC enzymatic phenotype. In some variations, it will be possible to use shuffled encoding sequences which have less than 21 contiguous nucleotides identical to a naturally-occurring PEPC gene sequence.

The invention provides shuffled PEPC encoding sequences, wherein the shuffled sequences comprise portions of a first parental PEPC encoding sequence

which comprises at least one mutation in the encoding sequence as compared to the collection of predetermined naturally occurring PEPC sequences.

Generally, the nomenclature used hereafter and the laboratory procedures in cell culture, molecular genetics, virology, and nucleic acid chemistry and hybridization described below are those well known and commonly employed in the art. Standard techniques are used for recombinant nucleic acid methods, polynucleotide synthesis, and microbial culture and transformation (e.g., biolistics, Agrobacterium (Ti plasmid), electroporation, lipofection). Generally enzymatic reactions and purification steps are performed according to the manufacturer's specifications. The techniques and procedures are generally performed according to conventional methods in the art and various general references (see, generally, Sambrook et al. Molecular Cloning: A Laboratory Manual, 2d ed. (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., which is incorporated herein by reference) which are provided throughout this document. The procedures therein are believed to be well known in the art and are provided for the convenience of the reader. All the information contained therein is incorporated herein by reference.

Oligonucleotides can be synthesized on an Applied Bio Systems oligonucleotide synthesizer according to specifications provided by the manufacturer.

Methods for PCR amplification are described in the art (PCR Technology: Principles and Applications for DNA Amplification ed. HA Erlich, Freeman Press, New York, NY (1992); PCR Protocols: A Guide to Methods and Applications, eds. Innis, Gelfland, Snisky, and White, Academic Press, San Diego, CA (1990); Mattila et al. (1991) Nucleic Acids Res. 19: 4967; Eckert, K.A. and Kunkel, T.A. (1991) PCR Methods and Applications 1: 17; PCR, eds. McPherson, Quirkes, and Taylor, IRL Press, Oxford; and U.S. Patent 4,683,202, which are incorporated herein by reference). Leaf PCR is suitable for genotype analysis of transgenote plants

All sequences referred to herein or equivalents which function in the disclosed methods can be retrieved by GenBank database file designation or a

commonly used reference name which is indexed in GenBank or otherwise published are incorporated herein by reference and are publicly available.

Incorporation by Reference of Related Applications

The following co-pending patent applications and publications of the present inventors and co-workers are incorporated herein by reference for all purposes: U.S.S.N. 08/198,431, filed 17 February 1994, PCT/US95/02126 filed 17 February 1995, WO97/20078, U.S. Patent 5,605,793, U.S. Patent 5,358,665, U.S. Patent 5,270,170, U.S.S.N. 08/425,684 filed 18 April 1995, U.S.S.N. 08/537,874 filed 30 October 1995, U.S.S.N. 08/564,955 filed 30 November 1995, U.S.S.N. 08/621,859 filed 25 March 1996, PCT/US96/05480 filed 18 April 1996, U.S.S.N. 08/650,400 filed 20 May 1996, U.S.S.N. 08/675,502 filed 3 July 1996, U.S.S.N. 08/721,824 filed 27 September 1996, U.S.S.N. 08/722,660 filed 27 September 1996, and U.S.S.N. 08/769,062 filed 18 December 1996; WO98/13485 and WO98/13487; and Stemmer (1995) Science 270: 1510; Stemmer et al. (1995) Gene 164: 49-53; Stemmer (1995) Bio/Technology 13: 549-553; Stemmer (1994) PNAS 91: 10747-10751; Stemmer (1994) Nature 370: 389-391; Crameri et al. (1996) Nature Medicine 2: 1-3; Crameri et al. (1996) Nature Biotechnology 14: 315-319; commonly assigned U.S. Patent Application "MODIFIED ADP-GLUCOSE PYROPHOSPHORYLASE FOR IMPROVEMENT AND OPTIMIZATION OF PLANT PHENOTYPES," USSN 60/107,782, filed on 10 November 1998 (Attorney docket number 018097-029000US); commonly assigned U.S. Patent Application U.S.S.N. 60/107,756 and 60/153,093 entitled "MODIFIED RIBULOSE BISPHOSPHATE CARBOXYLASE/OXYGENASE FOR IMPROVEMENT AND OPTIMIZATION OF PLANT PHENOTYPES," filed on 10 November 1998 and September 9, 1999, respectively; and "TRANSFORMATION, SELECTION, AND SCREENING OF SEQUENCE SHUFFLED POLYNUCLEOTIDES FOR DEVELOPMENT AND OPTIMIZATION OF PLANT PHENOTYPES" USSN 60/098,528, PCT/US99/19732 and USSN 09/385,833 filed August 31, 1998, August 30, 1999 and August 30, 1999, respectively.

Overview

The invention relates in part to a method for generating novel or improved PEPC genetic sequences and improved starch production phenotypes which do not naturally occur or would be anticipated to occur at a substantial frequency in nature. A broad aspect of the method employs recursive nucleotide sequence recombination, termed "sequence shuffling", which enables the rapid generation of a collection of broadly diverse phenotypes that can be selectively bred for a broader range of novel phenotypes or more extreme phenotypes than would otherwise occur by natural evolution in the same time period. A basic variation of the method is a recursive process comprising: (1) sequence shuffling of a plurality of species of a genetic sequence, which species may differ by as little as a single nucleotide difference or may be substantially different yet retain sufficient regions of sequence similarity or site-specific recombination junction sites to support shuffling recombination, (2) selection of the resultant shuffled genetic sequence to isolate or enrich a plurality of shuffled genetic sequences having a desired phenotype(s), and (3) repeating steps (1) and (2) on the plurality of shuffled genetic sequences having the desired phenotype(s) until one or more variant genetic sequences encoding a sufficiently optimized desired phenotype is obtained. In this general manner, the method facilitates the "forced evolution" of a novel or improved genetic sequence to encode a desired PEPC enzymatic phenotype which natural selection and evolution has heretofore not generated in the reference agricultural organism.

Typically, a plurality of PEPC genetic sequences are shuffled and selected by the present method. The method can be used with a plurality of alleles, homologs, or cognate genes of a gentic locus, or even with a plurality or genetic sequences from related organisms, and in some instances with unrelated genetic sequences or portions thereof which have recombinogenic portions (either naturally or generated via genetic engineering). Furthermore, the method can be used to evolve a heterologous PEPC sequence (e.g., a non-naturally occurring mutant gene from another species) to optimize its function and/or in a particular host cell.

PEPC

Coding sequences for various species are disclosed in the literature and Genbank, among other public sources, and may be obtained by cloning, PCR, or from deposited materials.

PEPC shufflants are generated by any suitable shuffling method from one or more parental sequences, optionally including mutagenesis, and the resultant shufflants are introduced into a suitable host cell, typically in the form of expression cassettes wherein the shuffled polynucleotide sequence encoding the PEPC is operably linked to a transcriptional regulatory sequence and any necessary sequences for ensuring transcription, translation, and processing of the encoded PEPC protein. Each such expression cassette or its shuffled PEPC encoding sequence can be referred to as a "library member" composing a library of shuffled PEPC sequences. The library is introduced into a population of host cells, such that individual host cells receive substantially one or a few species of library member(s), to form a population of shufflant host cells expressing a library of shuffled PEPC species. The population of shufflant host cells is screened so as to isolate or segregate host cells and/or their progeny which express PEPC having the desired enhanced phenotype. The shuffled PEPC encoding sequence(s) is/are recovered from the isolated or segregated shufflant host cells, and typically subjected to at least one subsequent round of mutagenesis and/or sequence shuffling, introduced into suitable host cells, and selected for the desired enhanced enzymatic phenotype; this cycle is generally performed iteratively until the shufflant host cells express a PEPC having the desired level or enzymatic phenotype or until the rate of improvement in the desired enzymatic phenotype produced by shuffling has substantially plateaued. The shufflant PEPC polynucleotides expressed in the host cells following the iterative process of shuffling and selection encode PEPC specie(s) having the desired enhanced phenotype.

For illustration and not to limit the invention, examples of a desired PEPC enzymatic phenotype can include increased substrate usage rate at a given substrate concentration, decreased inhibition by a PEPC inhibitor (desensitization), increased Km for inhibitor (desensitization), increased activation by an activator (desensitization), decreased Km for activator (desensitization), complete lack of need

for activation (desensitization), decreased ratio of Km for activator to Km for inhibitor, velocity (Vmax) for substrate use, desensitization to increased effects of inhibitor at increasing pH, and the like as described herein and as may be desired by the skilled artisan.

Shuffling

The following publications describe a variety of recursive recombination procedures and/or methods which can be incorporated into such procedures, e.g., for shuffling of PEPC genes and gene fragments as herein: Stemmer, et al., (1999) "Molecular breeding of viruses for targeting and other clinical properties. Tumor Targeting" 4:1-4; Nesset al. (1999) "DNA Shuffling of subgenomic sequences of subtilisin" Nature Biotechnology 17:893-896; Chang et al. (1999) "Evolution of a cytokine using DNA family shuffling" Nature Biotechnology 17:793-797; Minshull and Stemmer (1999) "Protein evolution by molecular breeding" Current Opinion in Chemical Biology 3:284-290; Christians et al. (1999) "Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling" Nature Biotechnology 17:259-264; Crameriet al. (1998) "DNA shuffling of a family of genes from diverse species accelerates directed evolution" Nature 391:288-291; Crameri et al. (1997) "Molecular evolution of an arsenate detoxification pathway by DNA shuffling," Nature Biotechnology 15:436-438; Zhang et al. (1997) "Directed evolution of an effective fucosidase from a galactosidase by DNA shuffling and screening" Proceedings of the National Academy of Sciences, U.S.A. 94:4504-4509; Patten et al. (1997) "Applications of DNA Shuffling to Pharmaceuticals and Vaccines" Current Opinion in Biotechnology 8:724-733; Crameri et al. (1996) "Construction and evolution of antibody-phage libraries by DNA shuffling" Nature Medicine 2:100-103; Crameri et al. (1996) "Improved green fluorescent protein by molecular evolution using DNA shuffling" Nature Biotechnology 14:315-319; Gates et al. (1996) "Affinity selective isolation of ligands from peptide libraries through display on a lac repressor 'headpiece dimer'" Journal of Molecular Biology 255:373-386; Stemmer (1996) "Sexual PCR and Assembly PCR" In: The Encyclopedia of Molecular Biology. VCH Publishers, New York. pp.447-457; Crameri and Stemmer (1995) "Combinatorial multiple cassette mutagenesis creates all the permutations of

mutant and wildtype cassettes" BioTechniques 18:194-195; Stemmer et al., (1995) "Single-step assembly of a gene and entire plasmid form large numbers of oligodeoxyribonucleotides" Gene, 164:49-53; Stemmer (1995) "The Evolution of Molecular Computation" Science 270: 1510; Stemmer (1995) "Searching Sequence Space" Bio/Technology 13:549-553; Stemmer (1994) "Rapid evolution of a protein in vitro by DNA shuffling" Nature 370:389-391; and Stemmer (1994) "DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution." Proceedings of the National Academy of Sciences, U.S.A. 91:10747-10751.

Additional details regarding DNA shuffling methods are found in U.S. Patents by the inventors and their co-workers, including: United States Patent 5,605,793 to Stemmer (February 25, 1997), "METHODS FOR IN VITRO RECOMBINATION;" United States Patent 5,811,238 to Stemmer et al. (September 22, 1998) "METHODS FOR GENERATING POLYNUCLEOTIDES HAVING DESIRED CHARACTERISTICS BY ITERATIVE SELECTION AND RECOMBINATION;" United States Patent 5,830,721 to Stemmer et al. (November 3, 1998), "DNA MUTAGENESIS BY RANDOM FRAGMENTATION AND REASSEMBLY;" United States Patent 5,834,252 to Stemmer, et al. (November 10, 1998) "END-COMPLEMENTARY POLYMERASE REACTION," and United States Patent 5,837,458 to Minshull, et al. (November 17, 1998), "METHODS AND COMPOSITIONS FOR CELLULAR AND METABOLIC ENGINEERING."

In addition, details and formats for DNA shuffling are found in a variety of PCT and foreign patent application publications, including: Stemmer and Crameri, "DNA MUTAGENESIS BY RANDOM FRAGMENTATION AND REASEMBLY" WO 95/22625; Stemmer and Lipschutz "END COMPLEMENTARY POLYMERASE CHAIN REACTION" WO 96/33207; Stemmer and Crameri "METHODS FOR GENERATING POLYNUCLEOTIDES HAVING DESIRED CHARACTERISTICS BY ITERATIVE SELECTION AND RECOMBINATION" WO 97/0078; Minshul and Stemmer, "METHODS AND COMPOSITIONS FOR CELLULAR AND METABOLIC ENGINEERING" WO 97/35966; Punnonen et al. "TARGETING OF GENETIC VACCINE VECTORS" WO 99/41402; Punnonen et

al. "ANTIGEN LIBRARY IMMUNIZATION" WO 99/41383; Punnonen et al. "GENETIC VACCINE VECTOR ENGINEERING" WO 99/41369; Punnonen et al. OPTIMIZATION OF IMMUNOMODULATORY PROPERTIES OF GENETIC VACCINES WO 9941368; Stemmer and Crameri, "DNA MUTAGENESIS BY RANDOM FRAGMENTATION AND REASSEMBLY" EP 0934999; Stemmer "EVOLVING CELLULAR DNA UPTAKE BY RECURSIVE SEQUENCE RECOMBINATION" EP 0932670; Stemmer et al., "MODIFICATION OF VIRUS TROPISM AND HOST RANGE BY VIRAL GENOME SHUFFLING" WO 9923107; Apt et al., "HUMAN PAPILLOMAVIRUS VECTORS" WO 9921979; Del Cardayre et al. "EVOLUTION OF WHOLE CELLS AND ORGANISMS BY RECURSIVE SEQUENCE RECOMBINATION" WO 9831837; Patten and Stemmer, "METHODS AND COMPOSITIONS FOR POLYPEPTIDE ENGINEERING" WO 9827230; Stemmer et al., and "METHODS FOR OPTIMIZATION OF GENE THERAPY BY RECURSIVE SEQUENCE SHUFFLING AND SELECTION" WO9813487.

Certain U.S. Applications provide additional details regarding DNA shuffling and related techniques, including "SHUFFLING OF CODON ALTERED GENES" by Patten et al. filed September 29, 1998, (USSN 60/102,362), January 29, 1999 (USSN 60/117,729), and September 28, 1999, USSN09/407,800 (Attorney Docket Number 20-28520US/PCT); "EVOLUTION OF WHOLE CELLS AND ORGANISMS BY RECURSIVE SEQUENCE RECOMBINATION", by del Cardyre et al. filed July 15, 1998 (USSN 09/166,188), and July 15, 1999 (USSN 09/354,922); "OLIGONUCLEOTIDE MEDIATED NUCLEIC ACID RECOMBINATION" by Crameri et al., filed February 5, 1999 (USSN 60/118,813) and filed June 24, 1999 (USSN 60/141,049) and filed September 28, 1999 (USSN 09/408,392, Attorney Docket Number 02-29620US); and "USE OF CODON-BASED OLIGONUCLEOTIDE SYNTHESIS FOR SYNTHETIC SHUFFLING" by Welch et al., filed September 28, 1999 (USSN 09/408,393, Attorney Docket Number 02-010070US); and "METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES & POLYPEPTIDES HAVING DESIRED CHARACTERISTICS" by Selifonov and Stemmer, filed February 5, 1999 (USSN

60/118854) and "METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES & POLYPEPTIDES HAVING DESIRED CHARACTERISTICS" by Selifonov et al. filed October 12, 1999 (USSN 09/416375).

As review of the foregoing publications, patents, published applications and U.S. patent applications reveals, recursive recombination and selection of nucleic acids to provide new nucleic acids with desired properties can be carried out by a number of established methods. Any of these methods can be adapted to the present invention to evolve PEPC coding nucleic acids or homologues to produce new enzymes with improved properties. Both the methods of making such enzymes and the enzymes or enzyme coding libraries produced by these methods are a feature of the invention.

In brief, at least 5 different general classes of recombination methods are applicable to the present invention. First, nucleic acids can be recombined in vitro by any of a variety of techniques discussed in the references above, including e.g., DNAse digestion of nucleic acids to be recombined followed by ligation and/or PCR reassembly of the nucleic acids. Second, nucleic acids can be recursively recombined in vivo, e.g., by allowing recombination to occur between nucleic acids in cells. Third, whole cell genome recombination methods can be used in which whole genomes of cells are recombined, optionally including spiking of the genomic or chloroplast recombination mixtures with desired library components such as PEPC encoding nucleic acids. Fourth, synthetic recombination methods can be used, in which oligonucleotides corresponding to different PEPC homologues are synthesized and reassembled in PCR or ligation reactions which include oligonucleotides which correspond to more than one parental nucleic acid, thereby generating new recombined nucleic acids. Oligonucleotides can be made by standard nucleotide addition methods, or can be made, e.g., by tri-nucleotide synthetic approaches. Fifth, in silico methods of recombination can be effected in which genetic algorithms are used in a computer to recombine sequence strings which correspond to PEPC homologues. The resulting recombined sequence strings are optionally converted into nucleic acids by synthesis of nucleic acids which correspond to the recombined

sequences, e.g., in concert with oligonucleotide synthesis/ gene reassembly techniques. Any of the preceding general recombination formats can be practiced in a reiterative fashion to generate a more diverse set of recombinant nucleic acids.

The above references provide these and other basic recombination formats as well as many modifications of these formats. Regardless of the format which is used, the nucleic acids of the invention can be recombined (with each other or with related (or even unrelated) nucleic acids to produce a diverse set of recombinant nucleic acids, including homologous nucleic acids.

Following recombination, any nucleic acids which are produced can be selected for a desired activity. A variety of related (or even unrelated) properties can be assayed for, using any available assay.

A basic format of the method, termed sequence shuffling (or simply "shuffling"), in broad application, consists of a method for generating a selected polynucleotide sequence or population of selected polynucleotide sequences, typically in the form of amplified and/or cloned polynucleotides, whereby the selected polynucleotide sequence(s) possess or encode a desired phenotypic characteristic (e.g., encode a polypeptide, promote transcription of linked polynucleotides, modify transformation efficiency, bind a protein, and the like) which can be selected for. One method of identifying polypeptides that possess a desired structure or functional property, such as encoding a desired enzymatic function(s) (e.g., an enhanced PEPC, a herbicide catabolizing enzyme, an optimized plant biosynthetic pathway), involves the screening of a large library of polynucleotides for individual library members which possess or encode the desired structure or functional property conferred by the polynucleotide sequence.

In a general aspect, the invention provides a method, termed "sequence shuffling", for generating libraries of recombinant polynucleotides having a desired PEPC enzyme characteristic which can be selected or screened for. Libraries of recombinant polynucleotides are generated from a population of related-sequence polynucleotides which comprise sequence regions which have substantial sequence identity and can be homologously recombined in vitro or in vivo. In the method, at least two species of the related-sequence polynucleotides are combined in a

recombination system suitable for generating sequence-recombined polynucleotides, wherein said sequence-recombined polynucleotides comprise a portion of at least one first species of a related-sequence polynucleotide with at least one adjacent portion of at least one second species of a related-sequence polynucleotide. Recombination systems suitable for generating sequence-recombined polynucleotides can be either: (1) in vitro systems for homologous recombination or sequence shuffling via amplification or other formats described herein, or (2) in vivo systems for homologous recombination or site-specific recombination as described herein. The population of sequence-recombined polynucleotides comprises a subpopulation of polynucleotides which possess desired or advantageous characteristics and which can be selected by a suitable selection or screening method. The selected sequence-recombined polynucleotides, which are typically related-sequence polynucleotides, can then be subjected to at least one recursive cycle wherein at least one selected sequence-recombined polynucleotide is combined with at least one distinct species of related-sequence polynucleotide (which may itself be a selected sequence-recombined polynucleotide) in a recombination system suitable for generating sequence-recombined polynucleotides, such that additional generations of sequence-recombined polynucleotide sequences are generated from the selected sequence-recombined polynucleotides obtained by the selection or screening method employed. In this manner, recursive sequence recombination generates library members which are sequence-recombined polynucleotides possessing desired characteristics. Such characteristics can be any property or attribute capable of being selected for or detected in a screening system, and may include properties of: an encoded protein, a transcriptional element, a sequence controlling transcription, RNA processing, RNA stability, chromatin conformation, translation, or other expression property of a gene or transgene, a replicative element, a protein-binding element, or the like, such as any feature which confers a selectable or detectable property.

Nucleic acid sequence shuffling is a method for recursive *in vitro* or *in vivo* homologous or nonhomologous recombination of pools of nucleic acid fragments or polynucleotides (e.g., genes from agricultural organisms or portions thereof). Mixtures of related nucleic acid sequences or polynucleotides are randomly or pseudo

randomly fragmented, and reassembled to yield a library or mixed population of recombinant nucleic acid molecules or polynucleotides.

The present invention is directed to a method for generating a selected polynucleotide sequence (e.g., a plant PEPC gene or microbe PEPC gene, or combinations thereof) or population of selected polynucleotide sequences, typically in the form of amplified and/or cloned polynucleotides, whereby the selected polynucleotide sequence(s) possess a desired phenotypic characteristic of PEPC enzymes which can be selected for, and whereby the selected polynucleotide sequences are genetic sequences having a desired functionality and/or conferring a desired phenotypic property to an agricultural organism in which the polynucleotide has been transferred into.

In a general aspect, the invention provides a method, called "sequence shuffling", for generating libraries of recombinant polynucleotides having a subpopulation of library members which encode an enhanced or improved PEPC protein. Libraries of recombinant polynucleotides are generated from a population of related-sequence PEPC polynucleotides which comprise sequence regions which have substantial sequence identity and can be homologously recombined in vitro or in vivo. In the method, at least two species of the related-sequence PEPC polynucleotides are combined in a recombination system suitable for generating sequence-recombined polynucleotides, wherein said sequence-recombined polynucleotides comprise a portion of at least one first species of a related-sequence PEPC polynucleotide with at least one adjacent portion of at least one second species of a related-sequence PEPC polynucleotide. Recombination systems suitable for generating sequence-recombined polynucleotides can be either: (1) in vitro systems for homologous recombination or sequence shuffling via amplification or other formats described herein, or (2) in vivo systems for homologous recombination or site-specific recombination as described herein, or template-switching of a retroviral genome replication event. The population of sequence-recombined polynucleotides comprises a subpopulation of PEPC polynucleotides which possess desired or advantageous enzymatic characteristics and which can be selected by a suitable selection or screening method. The selected sequence-recombined PEPC polynucleotides, which are typically

related-sequence polynucleotides, can then be subjected to at least one recursive cycle wherein at least one selected sequence-recombined PEPC polynucleotide is combined with at least one distinct species of related-sequence PEPC polynucleotide (which may itself be a selected sequence-recombined polynucleotide) in a recombination system suitable for generating sequence-recombined PEPC polynucleotides, such that additional generations of sequence-recombined polynucleotide sequences are generated from the selected sequence-recombined polynucleotides obtained by the selection or screening method employed. In this manner, recursive sequence recombination generates library members which are sequence-recombined polynucleotides possessing desired PEPC enzymatic characteristics. Such characteristics can be any property or attribute capable of being selected for or detected in a screening system.

Screening/selection produces a subpopulation of genetic sequences (or cells) expressing recombinant forms of PEPC gene(s) that have evolved toward acquisition of a desired enzymatic property. These recombinant forms can then be subjected to further rounds of recombination and screening/selection in any order. For example, a second round of screening/selection can be performed analogous to the first resulting in greater enrichment for genes having evolved toward acquisition of the desired enzymatic property. Optionally, the stringency of selection can be increased between rounds (e.g., if selecting for drug resistance, the concentration of drug in the media can be increased). Further rounds of recombination can also be performed by an analogous strategy to the first round generating further recombinant forms of the gene(s) or genome(s). Alternatively, further rounds of recombination can be performed by any of the other molecular breeding formats discussed. Eventually, a recombinant form of the PEPC gene(s) is generated that has fully acquired the desired enzymatic property.

In an embodiment, the first plurality of selected library members is fragmented and homologously recombined by PCR in vitro. Fragment generation is by nuclease digestion, partial extension PCR amplification, PCR stuttering, or other suitable fragmenting means, such as described herein and in WO95/22625 published 24 August 1995, and in commonly owned U.S.S.N. U.S.S.N. 08/621,859 filed 25

March 1996, PCT/US96/05480 filed 18 April 1996, which are incorporated herein by reference). Stuttering is fragmentation by incomplete polymerase extension of templates. A recombination format based on very short PCR extension times can be employed to create partial PCR products, which continue to extend off a different template in the next (and subsequent) cycle(s), and effect de facto fragmentation. Template-switching and other formats which accomplish sequence shuffling between a plurality of sequence-related polynucleotides can be used. Such alternative formats will be apparent to those skilled in the art.

In an embodiment, the first plurality of selected library members is fragmented in vitro, the resultant fragments transferred into a host cell or organism and homologously recombined to form shuffled library members in vivo.

In an embodiment, the first plurality of selected library members is cloned or amplified on episomally replicable vectors, a multiplicity of said vectors is transferred into a cell and homologously recombined to form shuffled library members in vivo.

In an embodiment, the first plurality of selected library members is not fragmented, but is cloned or amplified on an episomally replicable vector as a direct repeat or indirect (or inverted) repeat, which each repeat comprising a distinct species of selected library member sequence, said vector is transferred into a cell and homologously recombined by intra-vector or inter-vector recombination to form shuffled library members in vivo.

In an embodiment, combinations of in vitro and in vivo shuffling are provided to enhance combinatorial diversity. The recombination cycles (in vitro or in vivo) can be performed in any order desired by the practitioner.

In one embodiment, the first plurality of selected library members is fragmented and homologously recombined by PCR in vitro. Fragment generation is by nuclease digestion, partial extension PCR amplification, PCR stuttering, or other suitable fragmenting means, such as described herein and in the documents incorporated herein by reference. Stuttering is fragmentation by incomplete polymerase extension of templates.

In one embodiment, the first plurality of selected library members is fragmented in vitro, the resultant fragments transferred into a host cell or organism and homologously recombined to form shuffled library members in vivo. In an aspect, the host cell is a plant cell which has been engineered to contain enhanced recombination systems, such as an enhanced system for general homologous recombination (e.g., a plant expressing a recA protein or a plant recombinase from a transgene or plant virus) or a site-specific recombination system (e.g., a cre/LOX or frt/FLP system encoded on a transgene or plant virus).

In one embodiment, the first plurality of selected library members is cloned or amplified on episomally replicable vectors, a multiplicity of said vectors is transferred into a cell and homologously recombined to form shuffled library members in vivo in a plant cell, algae cell, or bacterial cell. Other cell types may be used, if desired.

In one embodiment, the first plurality of selected library members is not fragmented, but is cloned or amplified on an episomally replicable vector as a direct repeat or indirect (or inverted) repeat, which each repeat comprising a distinct species of selected library member sequence, said vector is transferred into a cell and homologously recombined by intra-vector or inter-vector recombination to form shuffled library members in vivo in a plant cell, algae cell, or microorganism.

In an embodiment, combinations of in vitro and in vivo shuffling are provided to enhance combinatorial diversity.

At least two additional related specific formats are useful in the practice of the present invention. The first, referred to as "in silico" shuffling utilizes computer algorithms to perform "virtual" shuffling using genetic operators in a computer. As applied to the present invention, PEPC sequence strings are recombined in a computer system and desirable products are made, e.g., by reassembly PCR or ligation of synthetic oligonucleotides, or other available techniques. In silico shuffling is described in detail in Selifonov and Stemmer in "METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES & POLYPEPTIDES HAVING DESIRED CHARACTERISTICS" filed 02/05/1999, USSN 60/118854 and "METHODS FOR MAKING CHARACTER STRINGS,

POLYNUCLEOTIDES & POLYPEPTIDES HAVING DESIRED CHARACTERISTICS" by Selifonov et al. filed October 12, 1999 (USSN 09/416375). In brief, genetic operators (algorithms which represent given genetic events such as point mutations, recombination of two strands of homologous nucleic acids, etc.) are used to model recombinational or mutational events which can occur in one or more nucleic acid, e.g., by aligning nucleic acid sequence strings (using standard alignment software, or by manual inspection and alignment) and predicting recombinational outcomes based upon selected genetic algorithms (mutation, recombination, etc.). The predicted recombinational outcomes are used to produce corresponding molecules, e.g., by oligonucleotide synthesis and reassembly PCR. As applied to the present invention, PEPC nucleic acids are aligned and recombined in silico, using any desired genetic operator, to produce character strings which are then generated synthetically for subsequent screening.

The second useful format is referred to as "oligonucleotide mediated shuffling" in which oligonucleotides corresponding to a family of related homologous nucleic acids (e.g., as applied to the present invention, families of PEPC variants) which are recombined to produce selectable nucleic acids. This format is described in detail in Crameri et al. "OLIGONUCLEOTIDE MEDIATED NUCLEIC ACID RECOMBINATION" filed February 5, 1999, USSN 60/118,813, Crameri et al. "OLIGONUCLEOTIDE MEDIATED NUCLEIC ACID RECOMBINATION" filed June 24, 1999, USSN 60/141,049; Crameri et al. "OLIGONUCLEOTIDE MEDIATED NUCLEIC ACID RECOMBINATION" filed September 28, 1999 (USSN 09/408,392, Attorney Docket Number 02-29620US); and "USE OF CODON-BASED OLIGONUCLEOTIDE SYNTHESIS FOR SYNTHETIC SHUFFLING" by Welch et al., filed September 28, 1999 (USSN 09/408,393, Attorney Docket Number 02-010070US). In brief, selected oligonucleotides corresponding to multiple homologous parental nucleic acids are synthesized, ligated and elongated (typically in a recursive format), typically either in a polymerase or ligase-mediated elongation reaction, to produce full-length PEPC nucleic acids. The technique can be used to recombine homologous or even non-homologous PEPC nucleic acid sequences.

One advantage of oligonucleotide-mediated recombination is the ability to recombine homologous nucleic acids with low sequence similarity, or even non-homologous nucleic acids. In these low-homology oligonucleotide shuffling methods, one or more set of fragmented nucleic acids (e.g., oligonucleotides corresponding to multiple PEPC nucleic acids) are recombined, e.g., with a set of crossover family diversity oligonucleotides. Each of these crossover oligonucleotides have a plurality of sequence diversity domains corresponding to a plurality of sequence diversity domains from homologous or non-homologous nucleic acids with low sequence similarity. The fragmented oligonucleotides, which are derived by comparison to one or more homologous or non-homologous nucleic acids, can hybridize to one or more region of the crossover oligos, facilitating recombination.

When recombining homologous nucleic acids, sets of overlapping family gene shuffling oligonucleotides (which are derived by comparison of homologous nucleic acids, by synthesis of corresponding oligonucleotides) are hybridized and elongated (e.g., by reassembly PCR or ligation), providing a population of recombined nucleic acids, which can be selected for a desired trait or property. The set of overlapping family shuffling gene oligonucleotides includes a plurality of oligonucleotide member types which have consensus region subsequences derived from a plurality of homologous target nucleic acids.

Typically, as applied to the present invention, family gene shuffling oligonucleotides which include one or more PEPC nucleic acid(s) are provided by aligning homologous nucleic acid sequences to select conserved regions of sequence identity and regions of sequence diversity. A plurality of family gene shuffling oligonucleotides are synthesized (serially or in parallel) which correspond to at least one region of sequence diversity.

Sets of fragments, or subsets of fragments used in oligonucleotide shuffling approaches can be provided by cleaving one or more homologous nucleic acids (e.g., with a DNASE), or, more commonly, by synthesizing a set of oligonucleotides corresponding to a plurality of regions of at least one nucleic acid (typically oligonucleotides corresponding to a full-length nucleic acid are provided as members of a set of nucleic acid fragments). In the shuffling procedures herein, these

cleavage fragments can be used in conjunction with family gene shuffling oligonucleotides, e.g., in one or more recombination reaction to produce recombinant PEPC nucleic acid(s).

One final synthetic variant worth noting is found in "SHUFFLING OF CODON ALTERED GENES" by Patten et al. filed September 29, 1998, (USSN 60/102,362), January 29, 1999 (USSN 60/117,729), and September 28, 1999, PCT/US99/22588 (Attorney Docket Number 20-28520US/PCT). As noted in detail in this set of related applications, one way of generating diversity in a set of nucleic acids to be shuffled (i.e., as applied to the present invention, PEPC nucleic acids), is to provide codon-altered nucleic acids which can be shuffled to provide access to sequence space not present in naturally occurring sequences. In brief, by synthesizing nucleic acids in which the codons which encode polypeptides are altered, it is possible to access a completely different mutational spectrum upon subsequent mutation of the nucleic acid. This increases the sequence diversity of the starting nucleic acids for shuffling protocols, which alters the rate and results of forced evolution procedures. Codon modification procedures can be used to modify any PEPC nucleic acid or shuffled nucleic acid, e.g., prior to performing DNA shuffling.

In brief, oligonucleotide sets comprising codon variations are synthesized and reassembled into full-length nucleic acids. The full length nucleic acids can themselves be shuffled (e.g., where the oligonucleotides to be reassembled provide sequence diversity at selected sites), and/or the full-length sequences can be shuffled by any available procedure to produce diverse sets of PEPC nucleic acids.

Improved Plants

Without reciting the various generalized formats of polynucleotide sequence shuffling and selection described previously or herein below, which will be referred to herein by the shorthand "shuffling", the present invention provides methods, compositions, and uses related to creating novel or improved plants, plant cells, algal cells, soil microbes, plant pathogens, commensal microbes, or other plant-related organisms having art-recognized importance to the agricultural, horticultural, and argonomic areas (collectively, "agricultural organisms"). In particular, any plant,

plant cell, algal cell, etc. can be transduced with a shuffled nucleic acid produced according to the present invention.

For example, agronomically and horticulturally important plant species can be transduced. Such species include, but are not restricted to, members of the families: *Graminae* (including corn, rye, triticale, barley, millet, rice, wheat, oats, etc.); *Leguminosae* (including pea, beans, lentil, peanut, yam bean, cowpeas, velvet beans, soybean, clover, alfalfa, lupine, vetch, lotus, sweet clover, wisteria, and sweetpea); *Compositae* (the largest family of vascular plants, including at least 1,000 genera, including important commercial crops such as sunflower) and *Rosaciae* (including raspberry, apricot, almond, peach, rose, etc.), as well as nut plants (including, walnut, pecan, hazelnut, etc.) Additionally, preferred targets for modification the evolved vectors of the invention, as well as those specified above, plants from the genera: *Agrostis, Allium, Antirrhinum, Apium, Arachis, Asparagus, Atropa, Avena* (e.g., oats), *Bambusa, Brassica, Bromus, Browaalia, Camellia, Cannabis, Capsicum, Cicer, Chenopodium, Chichorium, Citrus, Coffea, Coix, Cucumis, Curcubita, Cynodon, Dactylis, Datura, Daucus, Digitalis, Dioscorea, Elaeis, Eleusine, Festuca, Fragaria, Geranium, Glycine, Helianthus, Heterocallis, Hevea, Hordeum* (e.g., barley), *Hyoscyamus, Ipomoea, Lactuca, Lens, Lilium, Linum, Lolium, Lotus, Lycopersicon, Majorana, Malus, Mangifera, Manihot, Medicago, Nemesia, Nicotiana, Onobrychis, Oryza* (e.g., rice), *Panicum, Pelargonium, Pennisetum* (e.g., millet), *Petunia, Pisum, Phaseolus, Phleum, Poa, Prunus, Ranunculus, Raphanus, Ribes, Ricinus, Rubus, Saccharum, Salpiglossis, Secale* (e.g., rye), *Senecio, Setaria, Sinapis, Solanum, Sorghum, Stenotaphrum, Theobroma, Trifolium, Trigonella, Triticum* (e.g., wheat), *Vicia, Vigna, Vitis, Zea* (e.g., corn), the *Olyreae*, the *Pharoideae* and many others.

For example, common crop plants which are targets of the present invention include corn, rice, triticale, rye, cotton, soybean, sorghum, wheat, oats, barley, millet, sunflower, canola, peas, beans, lentils, peanuts, yam beans, cowpeas, velvet beans, clover, alfalfa, lupine, vetch, lotus, sweet clover, wisteria, sweetpea and nut plants (e.g., walnut, pecan, etc).

In certain variations, naturally occurring in vivo recombination mechanisms of plants, agricultural microorganisms, or vector-host cells for intermediate replication can be used in conjunction with a collection of shuffled polynucleotide sequence variants having a desired phenotypic property to be optimized further; in this way, a natural recombination mechanism can be combined with intelligent selection of variants in an iterative manner to produce optimized variants by "forced evolution", wherein the forced evolved variants are not expected to, nor are observed to, occur in nature, nor are predicted to occur at an appreciable frequency. The practitioner may further elect to supplement and/or the mutational drift by introducing intentionally mutated polynucleotide species suitable for shuffling, or portions thereof, into the pool of initial polynucleotide species and/or into the plurality of selected, shuffled polynucleotide species which are to be recombined. Mutational drift may also be supplemented by the use of mutagens (e.g., chemical mutagens or mutagenic irradiation), or by employing replication conditions which enhance the mutation rate.

Forced Evolution of Genes

The invention provides a means to evolve PEPC gene variants and/or suitable host cells, as well as providing a model system for evaluating a library of agents to identify candidate agents that could find use as agricultural reagents for commercial applications. Such agents may exhibit selectivity for inhibition of a naturally occurring PEPC enzyme and may be substantially less effective at inhibiting a shuffled PEPC enzyme which has been evolved to be resistant to the agent.

PEPC Shuffling Combinations

Although the skilled artisan may select alternative shuffling strategies for enhancing PEPC enzyme properties, the following general combinations can be used:

I. *Shuffling a PEPC gene from a first species of bacteria with a PEPC gene from a second species of bacteria.* The resultant shufflants may be transformed into bacterial host cells which preferably lack endogenous PEPC activity, algal cells, or plant cells for expression and selection. Phenotype selection of shufflants is typically performed by biochemical assay for PEPC, such as according to Gonzalez et al.

(1984) J. Plant Physiol. 116: 425; Devi et al. (1992) J. Plant Biochem. Biotech. 1: 73; Pairoba et al. (1996) Biosci. Biotech. Biochem. 60: 779; Salahas and Gavalas (1997) Photosynthetica 33: 189; or other suitable assay method selected by the artisan, or the like. Example bacteria for obtaining the PEPC gene(s) include *Rhodobacter sphaeroides*, *Rhodospirrilum rubrum*, *Escherichia coli*, *Salmonella typhimurium*, and the like. A preferred host cell is a strain of bacterium that is transformable and which lacks PEPC activity.

II. *Shuffling a parental plant PEPC encoding sequence with mutagenized variants thereof.* The resultant shufflants may be transformed into bacterial host cells which preferably lack endogenous plant-type PEPC activity (e.g., *E. coli*), algal cells, or plant cells for expression and selection. Phenotype selection of shufflants is typically performed by biochemical assay for PEPC activity or other suitable assay method selected by the artisan.

III. *Shuffling a PEPC from a first species of plant with a PEPC from a non-plant algae or bacterium, cyanobacteria.* The resultant shufflants may be transformed into host cells which preferably lack endogenous plant-type PEPC activity (e.g., *E. coli*), algal cells, or plant cells for expression and selection. Phenotype selection of shufflants is typically performed by biochemical assay for PEPC or other suitable assay method selected by the artisan. Example bacteria for the PEPC gene(s) include *Rhodobacter sphaeroides* (Falcone et al. (1998) J. Bact. 170: 5), *Rhodospirrilum rubrum* (Falcone and Tabita (1993) J.Bact. 175: 5066; Falcone et al. (1991) J. Bact. 173: 2099), *Escherichia coli*, *Salmonella typhimurium*, and the like. Example cyanobacteria that can serve as a source of PEPC genes include *Synechococcus, Cocochloris peniocystis, and Aphanizomenon flos-aquae*. Example green algae that can serve as sources of PEPC genes include *Euglena gracilis, Chlamadomonas reinhardii, and Anacystis nidulans*. Example plants that can serve as sources for the PEPC genes include corn, rice, maize, potato, wheat, rye, flax, cotton, pea, and the like.

IV. *Shuffling a plant PEPC from a first plant taxonomic species with a plant PEPC from a second plant taxonomic species.* The resultant shufflants may be transformed into host cells, which can preferably lack endogenous PEPC activity,

but which fold and process higher plant PEPC correctly for expression and selection. Phenotype selection of shufflants is typically performed by biochemical assay for PEPC or other suitable assay method selected by the artisan. Example higher plants that can serve as a source of PEPC genes include, but are not limited to: *Zea mays* (C4), *Amaranthus hybridus* (C4), *Glycine max* (C3), and *Nicotiana tabacum* (C3), among others.

V. *Shuffling a PEPC from a higher plant with mutagenized variants thereof.* A PEPC gene ("parental gene") from a species of C3 or C4 plant is subjected to mutagenesis and shuffling/selection to generate a population of mutagenized shufflants which have substantial sequence identity to the parental gene. The population of mutagenized shufflants is transferred into a population of host cells wherein the mutagenized shufflants are expressed and the resultant transformed host cell population is selected or screened for an enhanced PEPC phenotype. Phenotype selection of shufflants is typically performed by biochemical assay for PEPC activity or other suitable assay method selected by the artisan.

Transcriptional Regulatory Sequences

Suitable transcriptional regulatory sequences include: cauliflower mosaic virus 19S and 35S promoters, NOS promoter, OCS promoter, rbcS promoter, *Brassica* heat shock promoter, synthetic promoters, non-plant promoters modified, if necessary, for function in plant cells, substantially any promoter that naturally occurs in a plant genome, promoters of plant viruses or Ti plasmids, tissue-preferential promoters or cis-acting elements, light-responsive promoters or cis-acting elements (e.g., rbcS LRE), hormone-responsive cis-acting elements, developmental stage-specific promoters and cis-acting elements, viral promoters (e.g., from Tobacco Mosaic virus, Brome Mosaic Virus, Cauliflower Mosaic virus, and the like), and the like. In a variation, a transcriptional regulatory sequence from a first plant species is optimized for functionality in a second plant species by application of recursive sequence shuffling.

Transcriptional regulatory sequences for expression of shuffled PEPC sequences in chloroplasts is known in the art (Daniell et al. (1998) op.cit; O'Neill et

al. (1993) The Plant Journal 3: 729; Maliga P (1993) op.cit), as are homologous recombination vectors.

Host Cells for Screening PEPC Gene Shufflants

A variety of suitable host cells will be apparent to those skilled in the art. Of particular note, PEPC gene shufflants can be expressed in *E. coli*, as well as higher taxonomic host cells. However, PEPC from higher plants may not always be processed correctly in bacterial host cells, so higher plant PEPC gene shufflants may often be expressed for phenotype screening in plant cells, including mutant plant cell lines wherein an endogenous PEPC encoding gene has been functionally inactivated, preferably in homozygous format, to provide a plant cell substantially lacking endogenous PEPC activity, or the like.

Transformation

The transformation of plants and protoplasts in accordance with the invention may be carried out in essentially any of the various ways known to those skilled in the art of plant molecular biology. See, in general, Methods in Enzymology Vol. 153 ("Recombinant DNA Part D") 1987, Wu and Grossman Eds., Academic Press, incorporated herein by reference. Additional useful general references for plant cell cloning, culture and regeneration include Jones (ed) (1995) Plant Gene Transfer and Expression Protocols-- Methods in Molecular Biology, Volume 49 Humana Press Towata NJ; Payne et al. (1992) Plant Cell and Tissue Culture in Liquid Systems John Wiley & Sons, Inc. New York, NY (Payne); and Gamborg and Phillips (eds) (1995) Plant Cell, Tissue and Organ Culture; Fundamental Methods Springer Lab Manual, Springer-Verlag (Berlin Heidelberg New York) (Gamborg). A variety of cell culture media are described in Atlas and Parks (eds) The Handbook of Microbiological Media (1993) CRC Press, Boca Raton, FL (Atlas). Additional information for plant cell culture is found in available commercial literature such as the Life Science Research Cell Culture Catalogue (1998) from Sigma- Aldrich, Inc (St Louis, MO) (Sigma-LSRCCC) and, e.g., the Plant Culture Catalogue and supplement (1997) also from Sigma-Aldrich, Inc (St Louis, MO) (Sigma-PCCS). Additional details regarding plant cell culture are found in Croy, (ed.) (1993) Plant Molecular Biology Bios Scientific Publishers, Oxford, U.K. General texts discussing

cloning and other techniques relevant to the present invention, in a variety of contexts, include: Berger and Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology volume 152 Academic Press, Inc., San Diego, CA (Berger); Sambrook et al., Molecular Cloning - A Laboratory Manual (2nd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989 ("Sambrook") and Current Protocols in Molecular Biology, F.M. Ausubel et al., eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 1999) ("Ausubel")).

As used herein, the term transformation means alteration of the genotype of a host plant by the introduction of a nucleic acid sequence. The nucleic acid sequence need not necessarily originate from a different source, but it will, at some point, have been external to the cell into which it is to be introduced.

In one embodiment, the foreign nucleic acid is mechanically transferred by microinjection directly into plant cells by use of micropipettes. Alternatively, the foreign nucleic acid may be transferred into the plant cell by using polyethylene glycol. This forms a precipitation complex with the genetic material that is taken up by the cell (e.g., by incubation of protoplasts with "naked DNA" in the presence of polyethylenelycol)(Paszkowski et al., (1984) EMBO J. 3:2717-22; Baker et al (1985) Plant Genetics, 201-211; Li et al. (1990) Plant Molecular Biology Report 8(4)276-291].

In another embodiment of this invention, the introduced gene may be introduced into the plant cells by electroporation (Fromm et al., (1985) "Expression of Genes Transferred into Monocot and Dicot Plant Cells by Electroporation," Proc. Natl Acad. Sci. USA 82:5824, which is incorporated herein by reference). In this technique, plant protoplasts are electroporated in the presence of plasmids or nucleic acids containing the relevant genetic construct. Electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the plasmids. Electroporated plant protoplasts reform the cell wall, divide, and form a plant callus. Selection of the transformed plant cells with the transformed gene can be accomplished using phenotypic markers.

Cauliflower mosaic virus (CaMV) may also be used as a vector for introducing the foreign nucleic acid into plant cells (Hohn et al., (1982) "Molecular Biology of Plant Tumors," Academic Press, New York, pp.549-560; Howell, United States Patent No. 4,407,956). CaMV viral DNA genome is inserted into a parent bacterial plasmid creating a recombinant DNA molecule which can be propagated in bacteria. After cloning, the recombinant plasmid again may be cloned and further modified by introduction of the desired DNA sequence into the unique restriction site of the linker. The modified viral portion of the recombinant plasmid is then excised from the parent bacterial plasmid, and used to inoculate the plant cells or plants.

Another method of introduction of nucleic acid segments is high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface (Klein et al., (1987) Nature 327:70-73). Although typically only a single introduction of a new nucleic acid segment is required, this method particularly provides for multiple introductions.

A method of introducing the nucleic acid segments into plant cells is to infect a plant cell, an explant, a meristem or a seed with Agrobacterium tumefaciens transformed with the segment. Under appropriate conditions known in the art, the transformed plant cells are grown to form shoots, roots, and develop further into plants. The nucleic acid segments can be introduced into appropriate plant cells, for example, by means of the Ti plasmid of Agrobacterium tumefaciens. The Ti plasmid is transmitted to plant cells upon infection by Agrobacterium tumefaciens, and is stably integrated into the plant genome (Horsch et al., (1984) "Inheritance of Functional Foreign Genes in Plants," Science, 233:496-498; Fraley et al., (1983) Proc. Natl. Acad. Sci. USA 80:4803).

Ti plasmids contain two regions essential for the production of transformed cells. One of these, named transfer DNA (T DNA), induces tumor formation. The other, termed virulent region, is essential for the introduction of the T DNA into plants. The transfer DNA region, which transfers to the plant genome, can be increased in size by the insertion of the foreign nucleic acid sequence without its transferring ability being affected. By removing the tumor-causing genes so that they no longer interfere, the modified Ti plasmid can then be used as a vector for the

transfer of the gene constructs of the invention into an appropriate plant cell, such being a "disabled Ti vector."

All plant cells which can be transformed by Agrobacterium and whole plants regenerated from the transformed cells can also be transformed according to the invention so as to produce transformed whole plants which contain the transferred foreign nucleic acid sequence.

There are presently at least three different ways to transform plant cells with Agrobacterium: (1) co-cultivation of Agrobacterium with cultured isolated protoplasts; (2) transformation of cells or tissues with Agrobacterium, or (3) transformation of seeds, apices or meristems with Agrobacterium.

Method (1) uses an established culture system that allows culturing protoplasts and plant regeneration from cultured protoplasts.

Method (2) implies (a) that the plant cells or tissues can be transformed by Agrobacterium and (b) that the transformed cells or tissues can be induced to regenerate into whole plants.

Method (3) uses micropropagation. In the binary system, to have infection, two plasmids are needed: a T-DNA containing plasmid and a vir plasmid. Any one of a number of T-DNA containing plasmids can be used, the main issue being that one be able to select independently for each of the two plasmids.

After transformation of the plant cell or plant, those plant cells or plants transformed by the Ti plasmid so that the desired DNA segment is integrated can be selected by an appropriate phenotypic marker. These phenotypic markers include, but are not limited to, antibiotic resistance, herbicide resistance or visual observation. Other phenotypic markers are known in the art and may be used in this invention.

Protoplast Transformation

Numerous protocols for establishment of transformable protoplasts from a variety of plant types and subsequent transformation of the cultured protoplasts are available in the art and are incorporated herein by general reference. For examples, see Hashimoto et al. (1990) Plant Physiol. 93: 857; Plant Protoplasts, Fowke LC and Constabel F, eds., CRC Press (1994); Saunders et al. (1993)

Applications of Plant In Vitro Technology Symposium, UPM, 16-18 Nov. 1993; and Lyznik et al. (1991) BioTechniques 10: 295, each of which is incorporated herein by reference).

All plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be transformed by the present invention so that whole plants are recovered which contain the transferred foreign gene. Some suitable plants include, for example, species from the genera Fragaria, Lotus, Medicago, Onobrychis, Trifolium, Trigonella, Vigna, Citrus, Linum, Geranium, Manihot, Daucus, Arabidopsis, Brassica, Raphanus, Sinapis, Atropa, Capsicum, Hyoscyamus, Lycopersicon, Nicotiana, Solanum, Petunia, Digitalis, Majorana, Ciohorium, Helianthus, Lactuca, Bromus, Asparagus, Antirrhinum, Hererocallis, Nemesia, Pelargonium, Panicum, Pennisetum, Ranunculus, Senecio, Salpiglossis, Cucumis, Browaalia, Glycine, Lolium, Zea, Triticum, Sorghum, and Datura.

It is known that practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major cereal crop species, sugarcane, sugar beet, cotton, fruit and other trees, legumes and vegetables. Limited knowledge presently exists on whether all of these plants can be transformed by Agrobacterium. Species which are a natural plant host for Agrobacterium may be transformable in vitro. Although monocotyledonous plants, and in particular, cereals and grasses, are not natural hosts to Agrobacterium, work to transform them using Agrobacterium has also been successfully carried out by numerous investigators (Hooykas-Van Slogteren et al., (1984) Nature 311:763-764; Hernalsteens et al., (1984) EMBO J. 3:3039-41; Byteiber, et al. (1987) Proc. Natl. Acad. Sci. USA: 5345-5349; Graves and Goldman, (1986) Plant Mol. Biol 7: 43-50; Grimsley et al. (1988) Biochemistry 6: 185-189; WO 86/03776; Shimamoto et al. Nature (1989) 338: 274-276). Monocots may also be transformed by techniques or with vectors other than Agrobacterium. For example, monocots have been transformed by electroporation (Fromm et al. [1986] Nature 319:791-793; Rhodes et al. Science [1988] 240: 204-207), direct gene transfer (Baker et al. [1985] Plant Genetics 201-211), by using pollen-mediated vectors (EP 0 270 356), and by injection of DNA into floral tillers (de la Pena et al. [1987], Nature 325:274-276). Additional plant genera that may be transformed by Agrobacterium

include Chrysanthemum, Dianthus, Gerbera, Euphorbia, Pelaronium, Ipomoea, Passiflora, Cyclamen, Malus, Prunus, Rosa, Rubus, Populus, Santalum, Allium, Lilium, Narcissus, Ananas, Arachis, Phaseolus and Pisum.

Chloroplast Transformation

In certain embodiments, it may be desirable for the PEPC enzyme to be present in chloroplasts, possibly in combination with the more conventional cytosolic expression. As the PEPC enzyme of higher plants is encoded in the nuclear genome, it may be expressed with a fused chloroplast transit sequence peptide (CTS) to facilitate transloaction of the PEPC enzyme into chloroplasts, or it can be advantageous to transform the shufflant PEPC encoding sequences into chloroplasts if the host cells are derived from higher plants. Numerous methods are available in the art to accomplish the chloroplast transformation and expression (Daniell et al. (1998) op.cit; O'Neill et al. (1993) The Plant Journal 3: 729; Maliga P (1993) op.cit). The expression construct comprises a transcriptional regulatory sequence functional in plants operably linked to a polynucleotide encoding an enhanced PEPC protein. With respect to polynucleotide sequences encoding PEPC proteins, it may be desirable to express such encoding sequences in plastids, such as chloroplasts, for appropriate transcription, translation, and processing. With reference to expression cassettes which are designed to function in chloroplasts, such as an expression cassette encoding a PEPC in a higher plant, the expression cassette comprises the sequences necessary to ensure expression in chloroplasts - typically the encoding sequence is flanked by two regions of homology to the plastid genome so as to effect a homologous recombination with the chloroplastid genome; often a selectable marker gene is also present within the flanking plastid DNA sequences to facilitate selection of genetically stable transformed chloroplasts in the resultant transplastonic plant cells (see Maliga P (1993) TIBTECH 11: 101; Daniell et al. (1998) Nature Biotechnology 16: 346, and references cited therein).

Recovery of Selected Polynucleotide Sequences

A variety of selection and screening methods will be apparent to those skilled in the art, and will depend upon the particular phenotypic properties that are desired. The selected shuffled genetic sequences can be recovered for further

shuffling or for direct use by any applicable method, including but not limited to: recovery of DNA, RNA, or cDNA from cells (or PCR-amplified copies thereof) from cells or medium, recovery of sequences from host chromosomal DNA or PCR-amplified copies thereof, recovery of episome (e.g., expression vector) such as a plasmid, cosmid, viral vector, artificial chromosome, and the like, or other suitable recovery method known in the art.

Any suitable art-known method, including RT-PCR or PCR, can be used to obtain the selected shufflant sequence(s) for subsequent manipulation and shuffling.

Backcrossing

After a desired PEPC phenotype is acquired to a satisfactory extent by a selected shuffled gene or portion thereof, it is often desirable to remove mutations which are not essential or substantially important to retention of the desired phenotype ("superfluous mutations"). This is particularly desirable when the shuffled gene sequence is to be reintroduced back into a higher plant, as it is often preferred to harmonize the shufflant PEPC sequence with the endogenous PEPC sequence in the higher plant taxonomic species genome while retaining the desired PEPC phenotype obtained from the iterative shuffling/selection process. Superfluous mutations can be removed by backcrossing, which is shuffling the selected shuffled PEPC gene(s) with one or more parental PEPC gene and/or naturally-occurring PEPC gene(s) (or portions thereof) and selecting the resultant collection of shufflants for those species that retain the desired phenotype. By employing this method, typically in two or more recursive cycles of shuffling against parental or naturally-occurring PEPC genome(s) (or portions thereof) and selection for retention of the desired PEPC phenotype, it is possible to generate and isolate selected shufflants which incorporate substantially only those mutations necessary to confer the desired phenotype, whilst having the remainder of the genome (or portion thereof) consist of sequence which is substantially identical to the parental (or wild-type) sequence(s). As one example of backcrossing, a maize PEPC gene can be shuffled and selected for the capacity to substantially function in any Angiosperm plant cells; the resultant selected shufflants can be backcrossed with one or more PEPC genes of a particular plant species and

selected for the capacity to retain the capacity to confer the phenotype. After several cycles of such backcrossing, the backcrossing will yield gene(s) which contain the mutations necessary for the desired phenotype, and will otherwise have a genomic sequence substantially identical to the genome(s) of the host genome.

Isolated components (e.g., genes, regulatory sequences, replication origins, and the like) can be optimized and then backcrossed with parental sequences so as to obtain optimized components which are substantially free of superfluous mutations.

Transgenic Hosts

Transgenes and expression vectors to express shufflant PEPC sequences can be constructed by any suitable method known in the art; by either PCR or RT-PCR amplification from a suitable cell type or by ligating or amplifying a set of overlapping synthetic oligonucleotides; publicly available sequence databases and the literature can be used to select the polynucleotide sequence(s) to encode the specific protein desired, including any mutations, consensus sequence, or mutation kernel desired by the practitioner. The coding sequence(s) are operably linked to a transcriptional regulatory sequence and, if desired, an origin of replication. Antisense or sense-suppression transgenes and genetic sequences can be optimized or adapted for particular host cells and organisms by the described methods.

The transgene(s) and/or expression vectors are transferred into host cells, protoplasts, pluripotent embryonic plant cells, microbes, or fungi by a suitable method, such as for example lipofection, electroporation, microinjection, biolistics, *Agrobacterium tumefaciens* transduction of Ti plasmid, calcium phosphate precipitation, PEG-mediated DNA uptake, electroporation, electrofusion, or other method. Stable transfectant host cells can be prepared by art-known methods, as can transgenic cell lines.

Target Plants

As used herein, "plant" refers to either a whole plant, a plant part, a plant cell, or a group of plant cells. The class of plants which can be used in the method of the invention is generally as broad as the class of higher plants amenable to protoplast transformation techniques, including both monocotyledonous and

dicotyledonous plants. It includes plants of a variety of ploidy levels, including polyploid, diploid and haploid, and may employ non-regenerable cells for certain aspects which do not require development of an adult plant for selection or in vivo shuffling.

As noted, preferred plants for the transformation and expression of PEPC include agronomically and horticulturally important species. Such species include, but are not restricted to members of the families: *Graminae* (including corn, rye, triticale, barley, millet, rice, wheat, oats, etc.); *Leguminosae* (including pea, beans, lentil, peanut, yam bean, cowpeas, velvet beans, soybean, clover, alfalfa, lupine, vetch, lotus, sweet clover, wisteria, and sweetpea); *Compositae* (the largest family of vascular plants, including at least 1,000 genera, including important commercial crops such as sunflower) and *Rosaciae* (including raspberry, apricot, almond, peach, rose, etc.), as well as nut plants (including, walnut, pecan, hazelnut, etc.).

Targets for the invention also include plants from the genera: *Agrostis, Allium, Antirrhinum, Apium, Arachis, Asparagus, Atropa, Avena* (e.g., oats), *Bambusa, Brassica, Bromus, Browaalia, Camellia, Cannabis, Capsicum, Cicer, Chenopodium, Chichorium, Citrus, Coffea, Coix, Cucumis, Curcubita, Cynodon, Dactylis, Datura, Daucus, Digitalis, Dioscorea, Elaeis, Eleusine, Festuca, Fragaria, Geranium, Glycine, Helianthus, Heterocallis, Hevea, Hordeum* (e.g., barley), *Hyoscyamus, Ipomoea, Lactuca, Lens, Lilium, Linum, Lolium, Lotus, Lycopersicon, Majorana, Malus, Mangifera, Manihot, Medicago, Nemesia, Nicotiana, Onobrychis, Oryza* (e.g., rice), *Panicum, Pelargonium, Pennisetum* (e.g., millet), *Petunia, Pisum, Phaseolus, Phleum, Poa, Prunus, Ranunculus, Raphanus, Ribes, Ricinus, Rubus, Saccharum, Salpiglossis, Secale* (e.g., rye), *Senecio, Setaria, Sinapis, Solanum, Sorghum, Stenotaphrum, Theobroma, Trifolium, Trigonella, Triticum* (e.g., wheat), *Vicia, Vigna, Vitis, Zea* (e.g., corn), and the *Olyreae*, the *Pharoideae* and many others.

Common crop plants which are targets of the present invention include corn, rice, triticale, rye, cotton, soybean, sorghum, wheat, oats, barley, millet, sunflower, canola, peas, beans, lentils, peanuts, yam beans, cowpeas, velvet beans,

clover, alfalfa, lupine, vetch, lotus, sweet clover, wisteria, sweetpea and nut plants (e.g., walnut, pecan, etc).

Regeneration

Normally, regeneration will be involved in obtaining a whole plant from the transformation process. The term "transgenote" refers to the immediate product of the transformation process and to resultant whole transgenic plants.

The term "regeneration" as used herein, means growing a whole plant from a plant cell, a group of plant cells, a plant part or a plant piece (e.g. from a protoplast, callus, or tissue part).

Plant regeneration from cultural protoplasts is described in Evans et al., "Protoplasts Isolation and Culture," Handbook of Plant Cell Cultures 1:124-176 (MacMillan Publishing Co. New York 1983); M.R. Davey, "Recent Developments in the Culture and Regeneration of Plant Protoplasts," Protoplasts, (1983) - Lecture Proceedings, pp.12-29, (Birkhauser, Basal 1983); P.J. Dale, "Protoplast Culture and Plant Regeneration of Cereals and Other Recalcitrant Crops," Protoplasts (1983) - Lecture Proceedings, pp. 31-41, (Birkhauser, Basel 1983); and H. Binding, "Regeneration of Plants," Plant Protoplasts, pp.21-73, (CRC Press, Boca Raton 1985).

Additional details regarding plant regeneration are found in Jones (ed) (1995) Plant Gene Transfer and Expression Protocols-- Methods in Molecular Biology, Volume 49 Humana Press Towata NJ; Payne et al. (1992) Plant Cell and Tissue Culture in Liquid Systems John Wiley & Sons, Inc. New York, NY (Payne); Gamborg and Phillips (eds) (1995) Plant Cell, Tissue and Organ Culture; Fundamental Methods Springer Lab Manual, Springer-Verlag (Berlin Heidelberg New York) (Gamborg) and in Croy, (ed.) (1993) Plant Molecular Biology.

Regeneration from protoplasts varies from species to species of plants, but generally a suspension of transformed protoplasts containing copies of the exogenous sequence is first made. In certain species, embryo formation can then be induced from the protoplast suspension, to the stage of ripening and germination as natural embryos. The culture media will generally contain various amino acids and hormones, such as auxin and cytokinins. It is sometimes advantageous to add glutamic acid and proline to the medium, especially for such species as corn and

alfalfa. Shoots and roots normally develop simultaneously. Efficient regeneration will depend on the medium, on the genotype, and on the history of the culture. If these three variables are controlled, then regeneration is fully reproducible and repeatable.

Regeneration also occurs from plant callus, explants, organs or parts. Transformation can be performed in the context of organ or plant part regeneration. See, Methods in Enzymology, supra; also Methods in Enzymology, Vol. 118; and Klee et al., (1987) Annual Review of Plant Physiology, 38:467-486.

In vegetatively propagated crops, the mature transgenic plants are propagated by the taking of cuttings or by tissue culture techniques to produce multiple identical plants for trialling, such as testing for production characteristics. Selection of desirable transgenotes is made and new varieties are obtained thereby, and propagated vegetatively for commercial sale.

In seed propagated crops, the mature transgenic plants are self crossed to produce a homozygous inbred plant. The inbred plant produces seed containing the gene for the newly introduced foreign gene activity level. These seeds can be grown to produce plants that would produce the selected phenotype.

The inbreds according to this invention can be used to develop new hybrids. In this method a selected inbred line is crossed with another inbred line to produce the hybrid. The offspring resulting from the first experimental crossing of two parents is known in the art as the F1 hybrid, or first filial generation. Of the two parents crossed to produce F1 progeny according to the present invention, one or both parents can be transgenic plants.

Parts obtained from the regenerated plant, such as flowers, seeds, leaves, branches, fruit, and the like are covered by the invention, provided that these parts comprise cells which have been so transformed. Progeny and variants, and mutants of the regenerated plants are also included within the scope of this invention, provided that these parts comprise the introduced DNA sequences. Progeny and variants, and mutants of the regenerated plants are also included within the scope of this invention.

The following example is given to illustrate the invention, but are not to be limiting thereof.

EXPERIMENTAL EXAMPLE

EXAMPLE 1: Shuffling PEP Carboxylase

PEPC catalyzes the initial carbon fixation reaction in C4 plants such as maize and Sorghum, as well as Crassulacean acid metabolism (CAM) plants. There are other forms of PEPC involved in intermediary metabolism in all plants and microbes. PEPC involved in carbon fixation in C4 and CAM plants have been studied extensively with respect to its catalytic properties and regulation (Andero CS et al. (1987) FEBS Letters 213: 1; Chollet R (1996) Annu. Rev. Plant Physiol. Plant Mol. Biol. 47: 273).

cDNA coding for PEPC from various C4 and CAM plants are isolated using primers designed from published sequence in the gene bank (Devi M et al. (1992) op.cit; Chollet R (1996) op.cit and references therein). Complete coding sequence for PEPC can also be synthesized.

The PEPC genes from various related sources, which have high degree of homology at the nucleotide level are shuffled according to published procedures. Briefly, this procedure involves random fragmentation of the genes with DNAse I and selecting nucleotide fragments of 100-300 bp. The fragments are reassembled based on sequence similarity by primerless PCR. Recombination as well as variable levels of mutations that are introduced by the PCR reaction generate the diversity. The assembled genes can be cloned into *E. coli* or an *E. coli* mutant lacking PEPC. PEPC from C4 plants have been cloned and expressed in both prokaryotes and eukaryotes (Cretin et al. (1991) Gene 99: 87-94, Hudspath RL and Grula JW (1989) Plant Mol. Biol. 12: 579). Transformed colonies expressing a functional PEPC are screened by in vitro enzyme assay. Initial screening for expression of PEPC is also done using antibodies.

Colonies expressing shuffled PEPC genes can be selected and grown in larger amounts in liquid culture and assayed for specific properties. The assay procedure for PEPC involves coupling the activity with malic dehydrogenase and determining NADH disappearance spectrophotometrically at 340 nm (Gonzalez et al.

(1984) J. Plant Physiology 116: 425). The following properties are monitored in the shuffled PEPC by appropriate enzyme assays: (a) Activity at a broad pH range of 6-8.5 (b) desensitized to activation by various phosphorylated metabolites including glucose-6-phosphate (c) desensitized to feedback inhibitors malate and aspartate (d) other catalytic parameters such as Km for CO_2, phosphoenolpyruvate and Vmax.

PEPC shufflant genes from those clones expressing one or more of the desired properties mentioned above are iteratively shuffled in order to achieve optimization of each one of the properties mentioned above. The optimized PEPC gene, after appropriate modification for expression in plants, is used to transform the desired C4 crop in order to deregulate and increase carbon fixation.

Integrated Systems

The present invention provides computers, computer readable media and integrated systems comprising character strings corresponding to shuffled PEPC enzymes and corresponding enzyme-encoding nucleic acids. These sequences can be manipulated by in silico shuffling methods, or by standard sequence alignment or word processing software.

For example, different types of similarity and considerations of various stringency and character string length can be detected and recognized in the integrated systems herein. For example, many homology determination methods have been designed for comparative analysis of sequences of biopolymers, for spell-checking in word processing, and for data retrieval from various databases. With an understanding of double-helix pair-wise complement interactions among 4 principal nucleobases in natural polynucleotides, models that simulate annealing of complementary homologous polynucleotide strings can also be used as a foundation of sequence alignment or other operations typically performed on the character strings corresponding to the sequences herein (*e.g.*, word-processing manipulations, construction of figures comprising sequence or subsequence character strings, output tables, etc.). An example of a software package with algorithms for calculating sequence similarity is BLAST, which can be adapted to the present invention by inputting character strings corresponding to the sequences herein.

BLAST is described in Altschul *et al., J. Mol. Biol.* 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al., supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915).

An additional example of a useful sequence alignment algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments. It can also plot a tree showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng & Doolittle, *J. Mol. Evol.* 35:351-360 (1987). The method used is similar to the method described by Higgins & Sharp, *CABIOS* 5:151-153 (1989). The program can align, e.g., up to 300 sequences of a

maximum length of 5,000 letters. The multiple alignment procedure begins with the pairwise alignment of the two most similar sequences, producing a cluster of two aligned sequences. This cluster can then be aligned to the next most related sequence or cluster of aligned sequences. Two clusters of sequences can be aligned by a simple extension of the pairwise alignment of two individual sequences. The final alignment is achieved by a series of progressive, pairwise alignments. The program can also be used to plot a dendogram or tree representation of clustering relationships. The program is run by designating specific sequences and their amino acid or nucleotide coordinates for regions of sequence comparison.

The shuffled enzymes of the invention, or corresponding coding nucleic acids, are optinally sequenced and the sequences aligned to provide structure-function information. For example, the alignment of shuffled sequences which are selected for conversion activity against the same target provides an indication of which residues are relevant for conversion of the target (i.e., conserved residues are likely more important for activity than non-conserved residues).

Standard desktop applications such as word processing software (e.g., Microsoft Word™ or Corel WordPerfect™) and database software (e.g., spreadsheet software such as Microsoft Excel™, Corel Quattro Pro™, or database programs such as Microsoft Access™ or Paradox™) can be adapted to the present invention by inputting character strings corresponding to shuffled PEPC enzymes (or corresponding coding nucleic acids), e.g., shuffled by the methods herein. For example, the integrated systems can include the foregoing software having the appropriate character string information, e.g., used in conjunction with a user interface (e.g., a GUI in a standard operating system such as a Windows, Macintosh or LINUX system) to manipulate strings of characters. As noted, specialized alignment programs such as BLAST or PILEUP can also be incorporated into the systems of the invention for alignment of nucleic acids or proteins (or corresponding character strings).

Integrated systems for analysis in the present invention typically include a digital computer with software for aligning or manipulating sequences, as well as data sets entered into the software system comprising any of the sequences

herein. The computer can be, *e.g.*, a PC (Intel x86 or Pentium chip- compatible DOS™, OS2™ WINDOWS™ WINDOWS NT™, WINDOWS95™, WINDOWS98™ LINUX based machine, a MACINTOSH™, Power PC, or a UNIX based (*e.g.*, SUN™ work station) machine) or other commercially common computer which is known to one of skill. Software for aligning or otherwise manipulating sequences is available, or can easily be constructed by one of skill using a standard programming language such as Visual basic, Fortran, Basic, Java, or the like.

Any controller or computer optionally includes a monitor which is often a cathode ray tube ("CRT") display, a flat panel display (*e.g.*, active matrix liquid crystal display, liquid crystal display), or others. Computer circuitry is often placed in a box which includes numerous integrated circuit chips, such as a microprocessor, memory, interface circuits, and others. The box also optionally includes a hard disk drive, a floppy disk drive, a high capacity removable drive such as a writeable CD-ROM, and other common peripheral elements. Inputting devices such as a keyboard or mouse optionally provide for input from a user and for user selection of sequences to be compared or otherwise manipulated in the relevant computer system.

The computer typically includes appropriate software for receiving user instructions, either in the form of user input into a set parameter fields, *e.g.*, in a GUI, or in the form of preprogrammed instructions, *e.g.*, preprogrammed for a variety of different specific operations. The software then converts these instructions to appropriate language for instructing the system to carry out any desired operation.

In one aspect, the computer system is used to perform "in silico" shuffling of character strings. A variety of such methods are set forth in "METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES & POLYPEPTIDES HAVING DESIRED CHARACTERISTICS" by Selifonov and Stemmer, filed February 5, 1999 (USSN 60/118854) and "METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES & POLYPEPTIDES HAVING DESIRED CHARACTERISTICS" by Selifonov and Stemmer, filed October 12, 1999 (USSN 09/416,375). In brief, in the context of the present invention, genetic operators are used in genetic algorithms as described in the '375

application to change given ADPGPP sequences, e.g., by mimicking genetic events such as mutation, recombination, death and the like. Multi-dimensional analysis to optimize sequences can be also be performed in the computer system, e.g., as described in the '375 application.

A digital system can also instruct an oligonucleotide synthesizer to synthesize oligonucleotides, e.g., used for gene reconstruction or recombination, or to order oligonucleotides from commercial sources (e.g., by printing appropriate order forms or by linking to an order form on the internet).

The digital system can also include output elements for controlling nucleic acid synthesis (*e.g.*, based upon a sequence or an alignment of a shuffled enzyme as herein), i.e., an integrated system of the invention optionally includes an oligonucleotide synthesizer or an oligonucleotide synthesis controller. The system can include other operations which occur downstream from an alignment or other operation performed using a character string corresponding to a sequence herein, e.g., as noted above with reference to assays.

Combination Shuffling

One aspect of the present invention is the combinatorial shuffling of PEPC with other enzymes that affect carbon fixation. For example, one aspect of the present invention involves separately or simultaneously shuffling PEPC in combination with carbon fixation enzymes such as ribulose 1,5-bisphosphate carboxylase/oxygenase ("Rubisco"; EC 4.1.1.39), or with any Calvin cycle enzyme or Krebs cycle enzyme. Considerable detail regarding Rubisco and Calvin and Krebs cycle enzymes and shuffling of such enzymes to improve carbon fixation is found in commonly assigned U.S. Patent Application U.S.S.N. 60/107,756 and 60/153,093 entitled "MODIFIED RIBULOSE BISPHOSPHATE CARBOXYLASE/ OXYGENASE FOR IMPROVEMENT AND OPTIMIZATION OF PLANT PHENOTYPES," filed on 10 November 1998 and September 9, 1999, respectively and in "MODIFIED RIBULOSE BISPHOSPHATE CARBOXYLASE/ OXYGENASE FOR IMPROVEMENT AND OPTIMIZATION OF PLANT PHENOTYPES," by Stemmer et al., co-filed November 9, 1999 (Attorney Docket number 02-292-2US/PC). Shuffled PEPC genes and shuffled Rubisco genes are

optionally co-expressed in a cell or organism such as a plant to increase carbon fixation.

Similarly, shuffled Rubisco and shuffled ADP-glucose pyrophosphorylase ("ADPGPP"; EC 2.7.7.27; an enzyme involved in starch biosynthesis, e.g., in plants) can be expressed together in cells or plants to increase carbon fixation or to improve starch biosynthesis. Extensive details regarding ADP-glucose pyrophosphorylase gene shuffling are found in commonly assigned U.S. Patent Application U.S.S.N. 60/107,782, entitled "MODIFIED ADP-GLUCOSE PYROPHOSPHORYLASE FOR IMPROVEMENT AND OPTIMIZATION OF PLANT PHENOTYPES" filed on 10 November 1998 (Attorney docket number 018097-029000US) and co-filed application "MODIFIED ADP-GLUCOSE PYROPHOSPHORYLASE FOR IMPROVEMENT AND OPTIMIZATION OF PLANT PHENOTYPES" filed on 10 November 1999 (Attorney docket number 02-0290-1US). Of course, shuffled Rubisco, ADPGPP, and PEPC can all be expressed in a cell or organism such as a plant to increase carbon fixation, starch production, or the like.

In a further aspect, the present invention provides for the use of any apparatus, apparatus component, composition or kit herein, for the practice of any method or assay herein, and/or for the use of any apparatus or kit to practice any assay or method herein.

The foregoing description of the preferred embodiments of the present invention has been presented for purposes of illustration and description. They are not intended to be exhaustive or to limit the invention to the precise form disclosed, and many modifications and variations are possible in light of the above teaching.

Such modifications and variations which may be apparent to a person skilled in the art are intended to be within the scope of this invention.

All publications and patent applications herein are incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

WHAT IS CLAIMED IS:

1. A method for obtaining an isolated polynucleotide encoding an enhanced PEPC protein having PEPC catalytic activity wherein the PEPC enzymatic phenotype is significantly different than a protein encoded by a parental polynucleotide encoding a naturally-occurring PEPC enzyme, the method comprising:

recombining sequences of a plurality of parental polynucleotide species encoding at least one PEPC sequence under conditions suitable for sequence shuffling to form a resultant library of sequence-shuffled PEPC polynucleotides;

transferring said library into a plurality of host cells forming a library of transformants wherein sequence-shuffled PEPC polynucleotides are expressed;

assaying individual or pooled transformants for PEPC catalytic activity to determine the relative or absolute PEPC enzymatic phenotype and isolating a transformant having a PEPC enzymatic phenotype significantly different than parental PEPC, thereby identifying at least one enhanced transformant that expresses a PEPC enzyme activity which has a significantly altered compared to the PEPC activity encoded by the parental sequence(s);

recovering the sequence-shuffled PEPC polynucleotide from at least one enhanced transformant.

2. The method of claim 1, further comprising the step of subjecting a recovered sequence-shuffled PEPC polynucleotide encoding an enhanced PEPC to at least one subsequent round of recursive shuffling and selection, wherein said recovered sequence-shuffled PEPC polynucleotide is used as at least one parental sequence for subsequent shuffling.

3. The method of claim 1, wherein selection comprises assaying individual or pooled transformants for PEPC catalytic activity to determine the relative or absolute Km for substrate and identifying at least one enhanced transformant that expresses a PEPC activity which has a significantly lower Km for substrate than the PEPC activity encoded by the parental sequence(s).

4. The method of claim 1, wherein selection comprises assaying individual or pooled transformants for PEPC catalytic activity to determine the relative or absolute Km for inhibitor thereby identifying at least one enhanced transformant that expresses an PEPC activity which has a significantly higher Km for inhibitor than the PEPC activity encoded by the parental sequence(s).

5. The method of claim 1, wherein selection comprises assaying individual or pooled transformants for PEPC catalytic activity to determine the relative or absolute Km for activator thereby identifying at least one enhanced transformant that expresses an PEPC activity which has a significantly lower Km for activator than the PEPC activity encoded by the parental sequence(s).

6. The method of claim 1, wherein selection comprises assaying samples of individual transformants and their clonal progeny which are isolated into discrete reaction vessels for PEPC activity assay, or are assayed in situ.

7. The method of claim 1, wherein the host cell comprises a non-photosynthetic bacterium lacking an endogenous plant PEPC activity and is transformed with an expression cassette encoding a shufflant plant PEPC protein.

8. The method of claim 7, wherein the host cells harbor expression cassettes encoding a heterologous Rubisco or a heterologous PEPC.

9. The method of claim 1, wherein the plurality of host cells are plant cells.

10. The method of claim 1, wherein the plurality of host cells are plant cells, wherein the method further comprises regenerating transgenic plants from the host cells.

11. A plant cell protoplast and clonal progeny thereof containing a

sequence-shuffled polynucleotide encoding a PEPC which is not encoded by the naturally occurring genome of the plant cell protoplast.

12. A collection of plant cell protoplasts transformed with a library of sequence-shuffled PEPC polynucleotides in expressible form.

13. A regenerated plant containing at least one species of replicable or integrated polynucleotide comprising a sequence-shuffled portion and encoding an PEPC polypeptide.

14. A regenerated plant containing a polynucleotide expression cassette encoding a shuffled PEPC gene.

15. A regenerated plant of claim 13, further comprising a polynucleotide expression cassette encoding a shuffled bacterial or algal PEPC gene.

16. A polynucleotide encoding an enhanced PEPC protein having PEPC catalytic activity wherein: (1) the Km for substrate is significantly lower than a protein encoded by a parental polynucleotide encoding a naturally-occurring PEPC enzyme, (2) the Km for inhibitor is significantly higher than a protein encoded by a parental polynucleotide encoding a naturally-occurring PEPC enzyme, and/or (3) the Km for activator is significantly lower than a protein encoded by a parental polynucleotide encoding a naturally-occurring PEPC enzyme.

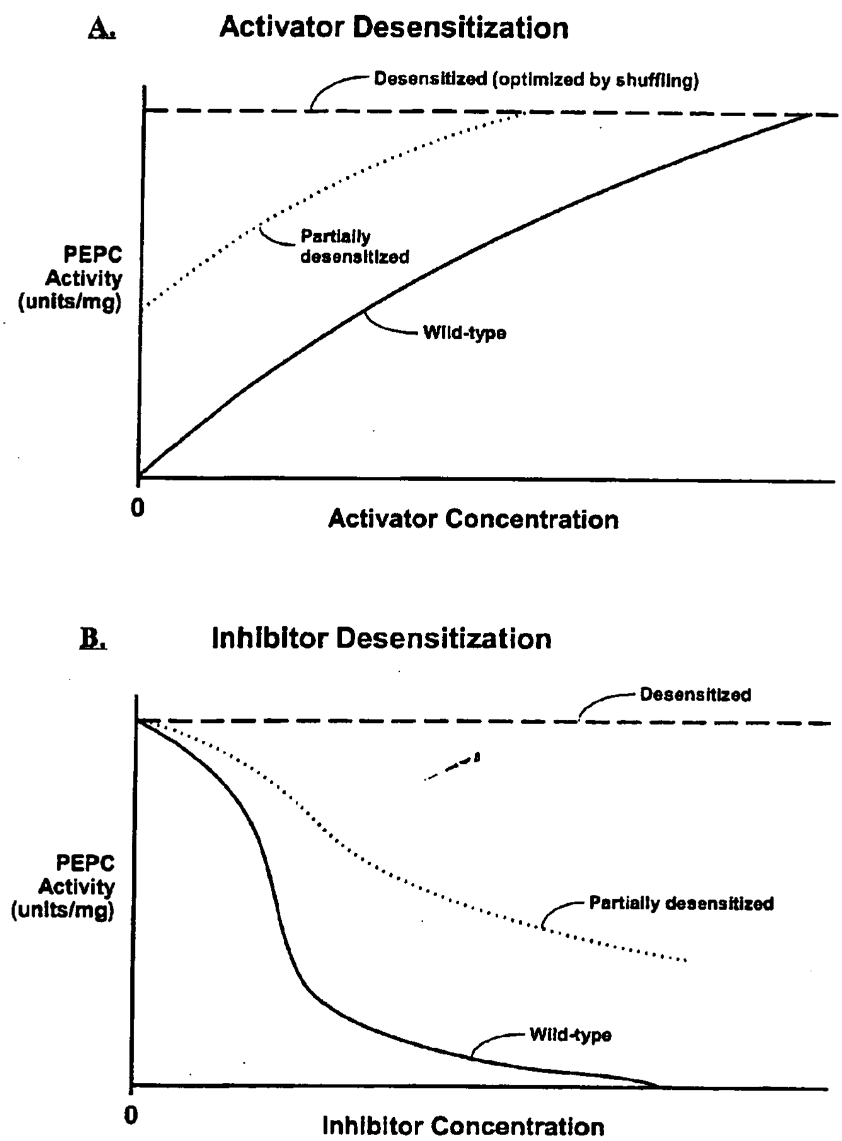
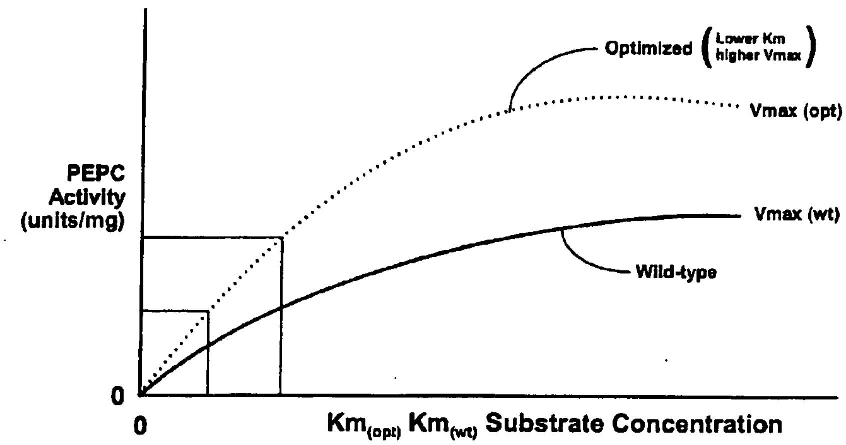

A.
Activator Desensitization
Desensitized (optimized by shuffling)
Partially desensitized
Wild-type
PEPC Activity (units/mg)
0
Activator Concentration
B.
Inhibitor Desensitization
Desensitized
Partially desensitized
Wild-type
PEPC Activity (units/mg)
0
Inhibitor Concentration

Figure 1

A. Substrate Optimization



B.

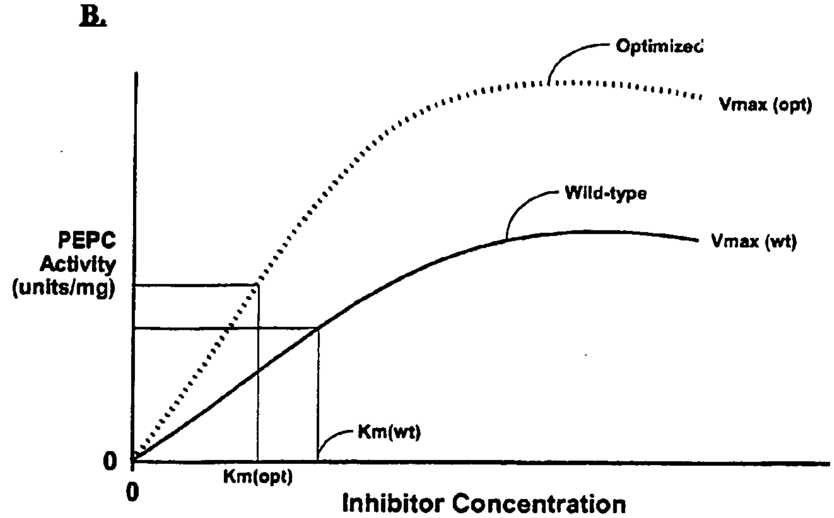


Figure 2

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 99/26771

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/10 C12Q1/527 C12Q1/68 C12N5/10 A01H5/00
C12N15/60

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CRAMERI A ET AL: "DNA SHUFFLING OF A FAMILY OF GENES FROM DIVERSE SPECIES ACCELERATESDIRECTED EVOLUTION" NATURE,GB,MACMILLAN JOURNALS LTD. LONDON, vol. 391, 15 January 1998 (1998-01-15), pages 288-291, XP000775869 ISSN: 0028-0836 the whole document	1-15

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
21 March 2000	05/04/2000

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Bilang, J

2

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 99/26771

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
Y	STEMMER W: "DNA shuffling by random fragmentatio and reassembly: In vitro recombination for molecular evolution" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA,US,NATIONAL ACADEMY OF SCIENCE. WASHINGTON, vol. 91, 1 October 1994 (1994-10-01), pages 10747-10751, XP002087463 ISSN: 0027-8424 the whole document figure 5	1-15
Y	CHOLLET R, ET AL.: "PHOSPHOENOLPYRUVATE CARBOXYLASE: A UBIQUITOUS, HIGHLY REGULATED ENZYME IN PLANTS" ANNUAL REVIEW OF PLANT PHYSIOLOGY AND PLANT MOLECULAR BIOLOGY, vol. 47, 1996, pages 273-298, XP000882055 cited in the application the whole document	1-15
A	WO 98 41622 A (NOVONORDISK AS) 24 September 1998 (1998-09-24) the whole document	1-15
A	HARAYAMA S: "Artificial evolution by DNA shuffling" TRENDS IN BIOTECHNOLOGY,GB,ELSEVIER PUBLICATIONS, CAMBRIDGE, vol. 16, no. 2, 1 February 1998 (1998-02-01), pages 76-82, XP004107046 ISSN: 0167-7799 the whole document	1-15
A	MORIKAWA M ET AL: "STUDIES ON THE ALLOSTERIC PROPERTIES OF MUTATIONALLY ALTERED PHOSPHOENOLPYRUVATE CARBOXYLASES OF ESCHERICHIA COLI: DISCRIMINATION OF ALLOSTERIC SITES" JOURNAL OF BIOCHEMISTRY,JP,JAPANESE BIOCHEMISTRY SOCIETY, TOKYO, vol. 81, no. 5, 1 January 1977 (1977-01-01), pages 1473-1485, XP000568820 ISSN: 0021-924X the whole document	1-15

2

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 99/26771

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9841622 A	24-09-1998	AU 6611498 A	12-10-1998

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: ______________________

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.